College of Engineering, Mathematics and Physical Sciences

Non-contact vision-based deformation monitoring on bridge structures

Submitted by Yan Xu to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Engineering
in June 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

Abstract

Information on deformation is an important metric for bridge condition and performance assessment, e.g. identifying abnormal events, calibrating bridge models and estimating load carrying capacities, etc. However, accurate measurement of bridge deformation, especially for long-span bridges remains as a challenging task. The major aim of this research is to develop practical and cost-effective techniques for accurate deformation monitoring on bridge structures. Vision-based systems are taken as the study focus due to a few reasons: low cost, easy installation, desired sample rates, remote and distributed sensing, etc.

This research proposes an custom-developed vision-based system for bridge deformation monitoring. The system supports either consumer-grade or professional cameras and incorporates four advanced video tracking methods to adapt to different test situations. The sensing accuracy is firstly quantified in laboratory conditions. The working performance in field testing is evaluated on one short-span and one long-span bridge examples considering several influential factors i.e. long-range sensing, low-contrast target patterns, pattern changes and lighting changes. Through case studies, some suggestions about tracking method selection are summarised for field testing. Possible limitations of vision-based systems are illustrated as well.

To overcome observed limitations of vision-based systems, this research further proposes a mixed system combining cameras with accelerometers for accurate deformation measurement. To integrate displacement with acceleration data autonomously, a novel data fusion method based on Kalman filter and maximum likelihood estimation is proposed. Through field test validation, the method is effective for improving displacement accuracy and widening frequency bandwidth. The mixed system based on data fusion is implemented on field testing of a railway bridge considering undesired test conditions (e.g. low-contrast target patterns and camera shake). Analysis results indicate that the system offers higher accuracy than using a camera alone and is viable for bridge influence line estimation.

With considerable accuracy and resolution in time and frequency domains, the potential of vision-based measurement for vibration monitoring is investigated. The proposed vision-based system is applied on a cable-stayed footbridge for deck deformation and cable vibration measurement under pedestrian loading. Analysis results indicate that the measured data enables accurate estimation of modal

frequencies and could be used to investigate variations of modal frequencies under varying pedestrian loads. The vision-based system in this application is used for multi-point vibration measurement and provides results comparable to those obtained using an array of accelerometers.

Acknowledgements

First, I would like to express my sincere thanks to my supervisors Professor James Brownjohn and Dr Ki Koo, for their patience, guidance and continuous support. This work wouldn't have been possible without them.

I would like to thank the two examiners, Prof. Necati Catbas from University of Central Florida and Dr. Prakash Kripakaran (internal) for their time and feedback.

A large part of this work is based on field testing. I'd like to thank the following groups for permission to use their bridges and for assistance they provided: the Humber Bridge Board, the Tamar Bridge Board, the West Somerset Railway and the Bridge Section of Devon County Council. Also thanks to the following colleagues for the support in field testing: David Hester, James Bassitt, Mungo Morgan, Farhad Huseynov, Karen Faulkner, Vincent Ao and Jose Capilla.

Thanks to all the staff and students in Vibration Engineering Section. It has been my pleasure to work with such a lovely group of people. My special thanks go to Dr. David Hester for his help and invaluable suggestions during my PhD study. I would also like to thank College of Engineering, Mathematics and Physical Sciences for the full financial support.

I appreciate the friendship of my two house mates, Dali and Fanlin.

Last but not least, I would like to thank my parents, my sister, and my husband for their love and support.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background

Bridges provide vital links in transportation networks and must be managed in a manner that minimises risk to public safety and disruption to service (See Fig. 1.1). However, visual inspection is still common for bridge maintenance that can be subjective and prone to error. Therefore, reliable approaches for bridge condition and performance assessment are required to assist the decision-making and make the best use of limited maintenance budget.

Fig. 1.1 Forth Road Bridge closure on Dec 2015 for approximately three weeks due to a local crack in the bridge truss under the carriageway (BBC News, 2015a).

As demonstrated in Fig. 1.2, structural health monitoring (SHM) is aimed at providing valuable information about structural performance and characterisation of structural defects to the asset owners, especially for those civil infrastructures beyond the design life. Vibration-based modal tests are a common way for structural condition and serviceability assessment, providing a direct view about the structural stiffness, mass properties and their distributions (Brownjohn et al., 2015). As well as for validating designs of civil structures, modal parameters extracted from vibration data obtained in short-term or long-term measurements are widely believed to have potential for identifying changes in structural condition or 'damage'

(Carden and Fanning, 2004), such as monitoring scour around bridge foundations from modal frequency shifts due to stiffness loss associated with scour (Prendergast and Gavin, 2014; Prendergast et al., 2013). The sensitivity of these parameters to 'damage' depends on the nature of the 'damage', for example local deterioration of material e.g. due to corrosion may not be detectable against background effects of environmental variability, whereas boundary conditions are known to have a relatively strong effect, for example fixity of bridge supports (Brownjohn et al., 2003).

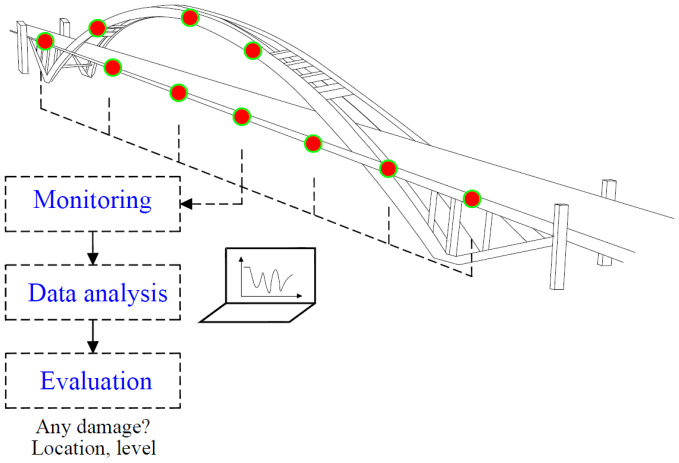


Fig. 1.2 Concept of Structural health monitoring.

Deformation is another important metric for bridge condition and performance assessment. For example, deflections measured under prescribed loads on aging bridges help to estimate load carrying capacities (BBC News, 2015b; Wang et al., 2011) that could assist the decision-making of bridge owners about the necessity of any strengthening work. Displacement is related to the structural stiffness, and extreme values might indicate either an extreme load or a deficiency in the structure. When recorded at high sample rates, displacement data provide valuable information about dynamic characteristics, and hence changes in structural condition. Therefore, accurate sensing approaches for bridge deformation monitoring are necessary and important.

For conventional displacement sensors such as linear variable differential transformers (LVDTs) and dial gauges, a stationary reference point is required that could be challenging in field tests. They are feasible only when the open space under bridge deck is accessible but require high installation efforts (Moreu et al., 2015). The indirect method through integrating the acceleration measurement is usually applied for short duration signals (e.g. a few seconds) and might fail to estimate static or quasi-static displacement components. Global positioning system (GPS) is commonly implemented for monitoring flexible bridges (e.g. long-span) since the range of their movement in operation is compatible with achievable GPS accuracy at sub-centimetre (Casciati and Fuggini, 2009) or centimetre level

(Nickitopoulou et al., 2006). Apart from accuracy (which differs from resolution), GPS performance is degraded during train passages (Moschas et al., 2013) and in cable-stayed or suspension bridges (Nickitopoulou et al., 2006)) due to multi-path noise.

Remote sensing techniques for displacement monitoring include robotic total stations (RTS), vision-based systems, laser Doppler vibrometers (LDV) and radar interferometry, etc. These sensors are easy to install with no dependence on a fixed reference point other than their own (stable) location. Access to test locations are sometimes still either necessary or recommended for the installation of assistant tools like reflective tapes for LDVs (Lou et al., 2017), reflective prisms for RTS (Brownjohn et al., 2015) and artificial targets for vision-based systems (Xu et al., 2016). Vision-based systems and radar interferometry both support multi-point simultaneous sensing while distributions of test locations for microwave interferometry systems are less flexible, depending on range resolution. Vision-based systems are the only type overcoming the dependence on expensive commercial products, and are thus receiving increased attention. Existing applications of vision-based systems for deformation monitoring in field tests cover a wide range of structural types including short-span (Chang and Xiao, 2010b; Feng et al., 2015a; Kim and Kim, 2011), long-span bridges (Kim and Kim, 2011; Macdonald et al., 1997; Stephen et al., 1993; Wahbeh et al., 2003; Ye et al., 2013), high-rise buildings (Liao et al., 2010) and stadium structures (Khuc and Catbas, 2017a).

1.2 Research objectives

This thesis seeks to develop practical and cost-effective techniques for accurate deformation monitoring on bridge structures. Vision-based systems are taken as the study focus due to a few reasons: low cost, easy installation, desired sample rates, remote and distributed sensing, etc.

There are three main goals of this study. The first goal is to develop an advanced vision-based system feasible for bridge deformation monitoring. The developed system has flexibility on video acquisition devices and offers the option of four advanced video tracking methods. A few laboratory and field monitoring tests are performed to investigate measurement accuracy and illustrate possible sensing challenges in field testing.

The second goal is to improve the working performance of the proposed vision-based system on site. To overcome observed limitations, a mixed system combining cameras with accelerometers (based on autonomous data fusion method) is proposed for stable and accurate deformation measurement.

The third goal is to validate the effectiveness of the proposed vision-based system for ambient modal testing on bridges. With considerable accuracy and

resolution in time and frequency domains, the potential of vision-based measurement for vibration monitoring is investigated through a field test on a cable-stayed footbridge.

1.3 Outline of thesis

The flowchart of chapter arrangement is presented in Fig. 1.3. The first part (Chapter 2 and 3) is the preliminary studies including the review about video processing methodologies and the trial tests using an expensive commercial vision-based system.

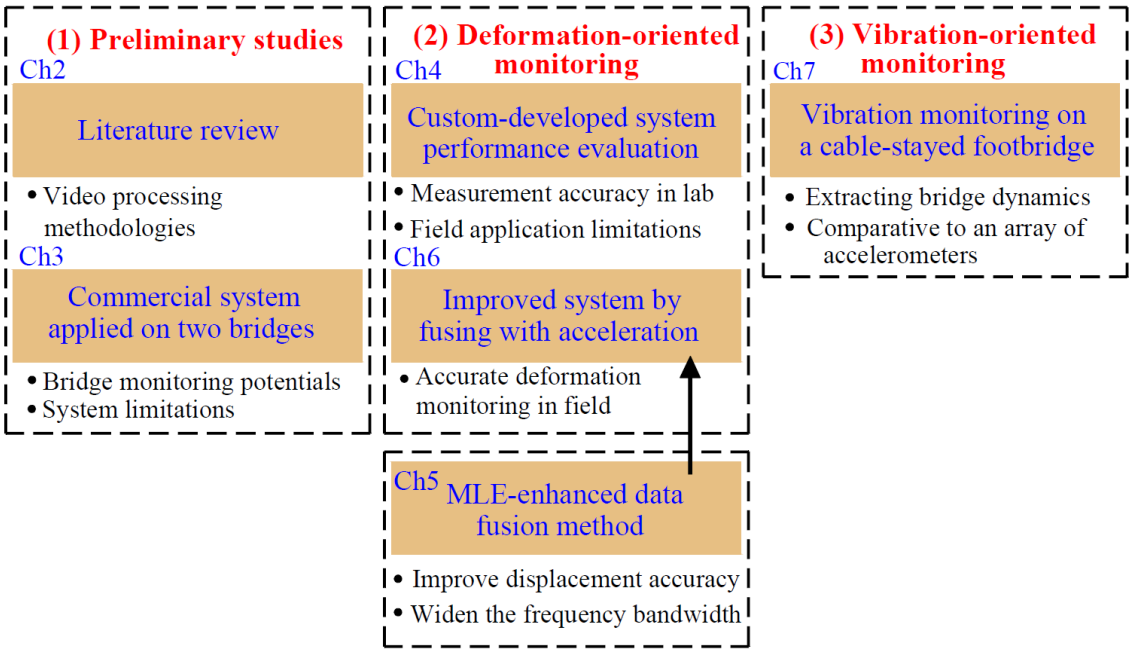


Fig. 1.3 Summary of thesis chapters. (Ch = Chapter)

Chapter 2 provides a literature review about methodologies and existing applications using vision-based systems for displacement monitoring in civil structures. Video processing methods are summarised with respect to principles, application examples, relative advantages and limitations. The two most active application fields (i.e. bridge deformation and cable vibration measurement) in literature are examined followed by a discussion of reported field challenges and limitations.

Chapter 3 introduces the application experience using a commercial vision-based system for deformation monitoring on two bridge structures. The purpose of these tests is to illustrate the potential and to build confidence in vision-based structural monitoring. Based on field experiences, a discussion about this commercial system is presented regarding installation efforts, application ranges and measurement performance.

With a sound understanding of video processing methodologies, the second part (Chapter 4-6) of the thesis is to develop an advanced vision-based system and to evaluate its working performance in field testing.

The proposed system is introduced in Chapter 4. The system has flexibility on video acquisition devices and offers four advanced video tracking methods. A detailed description about video processing methods is provided in Appendix A. Chapter 4 also investigates the working performance of the proposed system through three case studies (one in a laboratory, one on a short-span bridge and one on a long-span bridge). The tracking accuracy is quantified in laboratory conditions while measurement performance under several field challenges are investigated including long-range sensing, low-contrast target patterns, pattern changes and lighting changes.

To improve measurement accuracy, an effective way is data fusion which allows the integration of data from a series of sensors to make a better estimation than that from any sensor alone. Chapter 5 proposes a practical data fusion method to merge displacement data with the collocated acceleration for accuracy improvement. The main algorithm of data fusion method is based on multi-rate Kalman filter assisted with the maximum likelihood estimation (MLE) that is implemented to determine the necessary noise parameters automatically. The MLE enhanced data fusion method is validated on numerical and field data, indicating the capacity of improving displacement accuracy and widening frequency bandwidth.

In Chapter 6, a mixed system (i.e. cameras combined with accelerometers) for accurate deformation sensing is proposed based on the data fusion method and validated in field testing. The system considers automatic compensation of camera shake and autonomous data fusion, and the result is capable of achieving high accuracy even with relatively low-contrast target patterns. The system is demonstrated through a field monitoring test on a short-span railway bridge during train passages and is validated to offer higher accuracy than using a camera alone.

Since vision-based systems could measure at considerable sample rates, the potential for vibration monitoring is investigated in the third part (Chapter 7) .

In Chapter 7, the vision-based system is applied on a cable-stayed footbridge for deck deformation and cable vibration measurement under pedestrian loading. Analysis results indicate that the measured data enables accurate estimation of modal frequencies and could be used to investigate variations of modal frequencies under varying pedestrian loads. The vision-based system in this application is used for multi-point vibration measurement and provides results comparable to those obtained using an array of accelerometers.

Chapter 8 concludes the thesis and discusses future work.

Chapter 2

Review of vision-based methodologies

Vision-based systems are promising tools for displacement measurement in civil structures, receiving increased attention. Approximately one hundred papers to date have appeared on this subject, investigating topics like: system development and improvement, the viability on field applications and the potential for structural condition assessment. This chapter presents a literature review of vision-based displacement monitoring mainly from the perspectives of methodologies.

Video processing procedures in this paper are summarised as a three-component framework, camera calibration, target tracking and structural displacement calculation. Methods for each component are presented with respect to principle, relative advantages and limitations. Applications in the two most active fields: bridge deformation and cable vibration measurement are examined followed by a summary of field challenges observed in monitoring tests.

This chapter is reproduced from the paper titled "**Review of machine-vision based methodologies for displacement measurement in civil structures**", co-authored with James Brownjohn, which was published in *Journal of Civil Structural Health Monitoring* (Xu and Brownjohn, 2018a).

2.1 Introduction

Vision-based systems offer significant potential for structural condition assessment, in particular for system identification (Caetano et al., 2007; Oh et al., 2015; Yoon et al., 2016). In addition, deformation information has been used for finite element model calibration (Feng and Feng, 2015), damage detection (Cha et al., 2017) and contribution to bridge weigh-in-motion system with camera assistance for traffic monitoring (Ojio et al., 2016).

Investigations have been made in system improvement in both video acquisition hardware and video processing algorithms. The feasible video acquisition devices

are expanded to include smartphone cameras (Yoon et al., 2016; Zhao et al., 2017), while artificial targets required in conventional systems were discarded in some recent applications under specific camera configurations (Feng et al., 2015a; Khuc and Catbas, 2017a; Yoon et al., 2016). Efficient target tracking techniques in the computer vision field have been validated in structural deformation monitoring (Chen et al., 2015b; Khuc and Catbas, 2017a; Yoon et al., 2016) and the measurement results describing structural displacement have been expanded to three-dimensional (Chang and Ji, 2007; Oh et al., 2015; Santos et al., 2016; Schreier, 2004) and six degree of freedom (DOF) motions (Chang and Xiao, 2010b; Martins et al., 2015).

This chapter aims to present a summary of key work in the field of vision-based systems for structural displacement monitoring while highlighting the principles, advantages and shortcomings of these systems. Although previous reviews of vision-based structural monitoring exist (Baqersad et al., 2017; Wu and Casciati, 2014; Ye et al., 2016), the contribution of this work is to provide an overview of system classifications, methodologies and applications in field monitoring. It is organised as follows. The components of a vision-based system for displacement monitoring are introduced, followed by a comparison of several mature vision-based systems in application scopes in section 2.2. In section 2.3, vision-based systems are categorised based on methods of video processing, with three components in video processing procedures (i.e. camera calibration, target tracking and structural displacement calculation) reviewed in terms of principle, applications, advantages and shortcomings, respectively. In section 2.4, applications for bridge deformation and cable vibration measurement are reviewed followed by a discussion of measurement challenges in field applications. Finally, important gaps requiring further investigation are presented e.g. robust tracking methods, non-contact sensing and measurement accuracy evaluation in field conditions.

2.2 Vision-based displacement monitoring systems

Applying a vision-based system for structural displacement monitoring requires setting up one or more cameras in a stable location, looking at the ‘target’ contained in a structure and deriving the structural displacement through target tracking. Here the ‘target’ could be either artificial (e.g. pre-installed marker, LED lamp or planar panel with special patterns) or an existing structural feature (e.g. bolts or holes). As shown in Fig. 2.1, the hardware generally comprises camera, lens, laptop/portable computer with video-processing package and some accessories e.g. tripod. The video processing algorithm is critical: its role is acquiring the video frames covering the target region, tracking the target locations in image sequences and finally transforming the target location in image to time history of structural displacement.

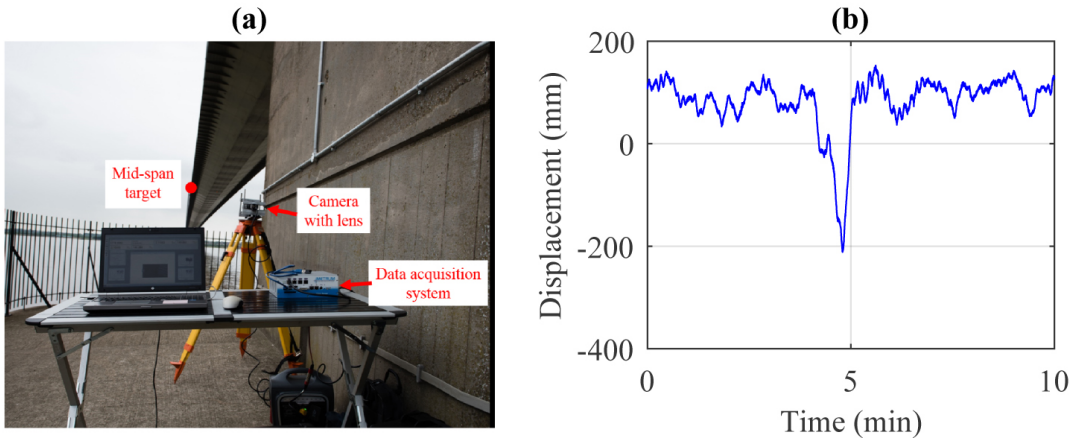


Fig. 2.1 Vision-based system for structural displacement monitoring of the Humber Bridge (Brownjohn et al., 2017): (a) site configuration of the vision-based monitoring system; and (b) 10-min time history signal of vertical displacement at the bridge mid-span measured by the vision-based monitoring system.

Systems for extracting metric information from images or videos exist in several fields as indicated in Table 2.1 e.g. digital image correlation (DIC) (Kim and Kim, 2011; Schreier et al., 2009; Yoneyama et al., 2007), photogrammetric techniques (Jáuregui et al., 2003) and motion capture systems (MCS) (Oh et al., 2015; Park et al., 2015). DIC is a measurement tool to extract full-field displacements or strains of a member surface in experimental solid mechanics (Pan et al., 2009; Schreier et al., 2009; Sutton et al., 2008). Photogrammetry, originally in the production of topographic maps (Hartley and Mundy, 1993), is expanded to include deflection monitoring of bridge structures (Jiang et al., 2008). Motion capture systems (MCS) are usually applied to capture the movements of a high degree-of-freedom skeleton structure with a number of joints (e.g. human bodies) (Moeslund and Granum, 2001).

A vision-based system for structural displacement monitoring owns its unique features, as indicated in the last row of Table 2.1. Researchers have performed several investigations into system development targeted at structural applications and these studies will be reviewed in terms of methodologies in the next section.

2.3 Review of vision-based structural displacement measurement

In this study, vision-based systems in literature are classified based on video-processing methodologies. A typical video processing package could fit into a three-component framework shown in Fig. 2.2. The derived displacement data could be interpreted for bridge condition assessment.

Table 2.1 Summary of vision-based systems

Vision-based systems	Main study objects	Measurement information	Features
Digital image correlation (DIC)	Small-scale experimental members under large distortional deformation	Full-field displacements or strains on member surface	Laboratory application in controlled environment; Fixed camera locations; Dense measurement with high resolution; Usually large deformation with shape distortion.
Motion capture systems (MCS)	Objects or human bodies with a high degree-of-freedom skeleton structure	3D locations of each joints in structure	Laboratory application in controlled environment; Fixed camera locations; At least two cameras with overlapped views; Markers and calibration object for calibration assistance.
Photogrammetry	Initially aerial and terrestrial applications; now bridges under live loads	3D geometry of objects and deflection measurement	Field applications on structures mainly in stationary status; Movable locations of camera; Distributed control points for calibration assistance.
System for structural monitoring	Structures with small deformation compared with structure scale	2D or 3D displacement with proper sample rate	Field applications and easy installation preferred; High accuracy and also high calculation efficiency (for real-time dynamic measurement); Small deformation compared with structure scale and camera-to-structure distance.

If the monitoring campaign is only for system identification and exact vibration values (Chen et al., 2015a; Kim and Kim, 2013) are not required, target tracking may be the only part of the whole video processing procedure needed, but coordinate transformation might be necessary to align the image motion directions with the structural axes. Next, the methods for camera calibration, target tracking and structural displacement calculation in literature are reviewed separately.

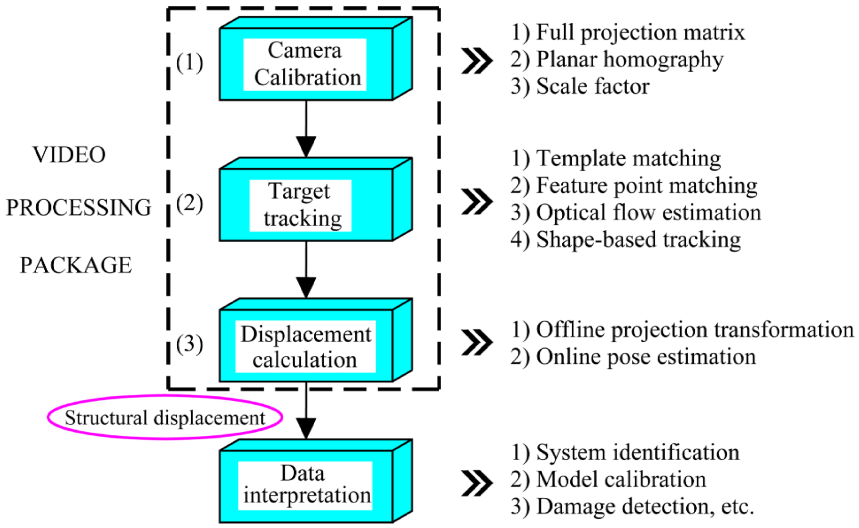


Fig. 2.2 Video processing procedures for structural displacement measurement and common methods in each step.

2.3.1 Camera calibration

Camera calibration concerns building the projection relationship between the 3D structural points in the structural coordinate system and the corresponding 2D points in the image plane. The determined projection transformation could be used to recover the real locations of targets in structure given the target locations in the image.

Three categories of projection transformation are reported in the literature including the full projection matrix, planar homography and scale factor as indicated in Table 2.2. In most cases, the projection transformation is following the full perspective model while it could be simplified to an affine camera model when cameras are equipped with large focal length lenses (Santos et al., 2016).

Table 2.2 Projection transformation from structure to image plane

Projection transformation		Assumptions	Recovered localisation information of target
(1)	Full projection matrix	–	3D structural coordinates
(2)	Planar homography	The motion along one axis in structural coordinate system is negligible	2D structural coordinates
(3)	Scale factor	The camera optical axis is perpendicular to one plane in the structural coordinate system (e.g. the target plane XY)	2D motions within the target plane

2.3.1.1 Full projection matrix

Principle The full projection matrix is the general form of projection transformation from the 3D structural system to the 2D image plane under no constraint on camera orientation and structural movement directions and is usually used to reconstruct the target 3D structural displacement. The projection relationship is demonstrated in Fig. 2.3 with a point $\mathbf{P}_s$ ($\mathbf{X}_w = [\text{X,Y,Z,1}]^T$) in the structural coordinate system mapping to a point $\mathbf{P}_I$ ($\mathbf{u} = [u, v, 1]^T$) in the 2D image plane,

$$\alpha\{\mathbf{u}\} = [\mathbf{H}]_{3\times4}\{\mathbf{X}_w\} \tag{2.1}$$

where $[\mathbf{H}]_{3\times4}$ is a full projection matrix and α is an arbitrary coefficient.

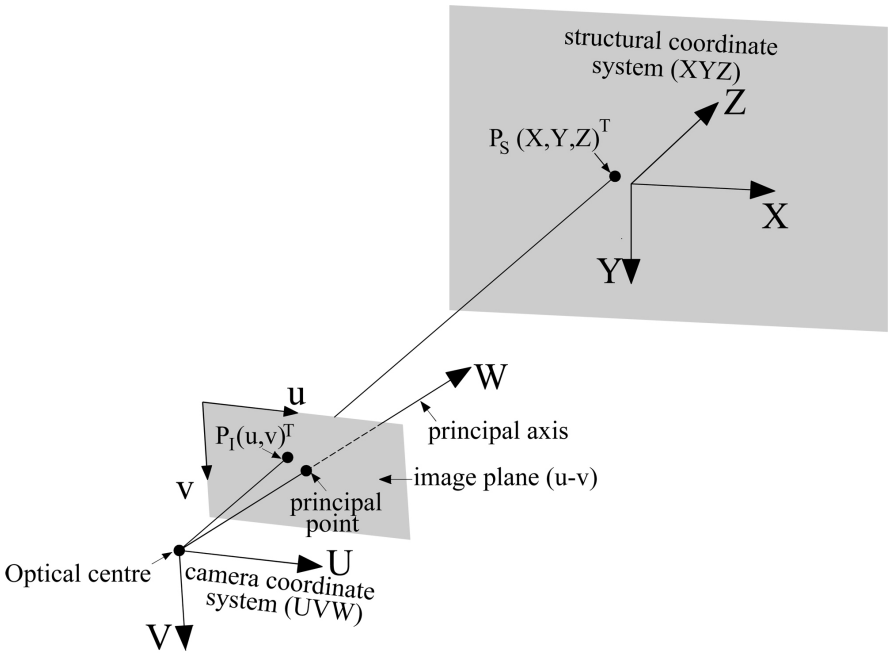


Fig. 2.3 Video processing procedures for structural displacement measurement and common methods in each step.

The calibration process is shown in Fig. 2.4 with two main steps. The camera intrinsic matrix is usually estimated in the laboratory by analysing a set of images of a calibration object taken from different viewpoints (Zhang, 2000). The calibration object is typically a flat plane or 3D object with a dot or grid pattern of known spacing such as the chessboard pattern shown in Fig. 2.4. At least three views of the calibration object with four corner points are required, but it is suggested to use at least ten images to derive more robust estimates (Bradski and Kaehler, 2008). After laboratory calibration, any lens functions e.g. autofocus and automated image stabilisation that might lead to changes in camera internal parameters are disabled.

Consumer-grade cameras and smartphone cameras always employ wide angle lenses to increase the field-of-view (Yoon et al., 2016), leading to distorted images

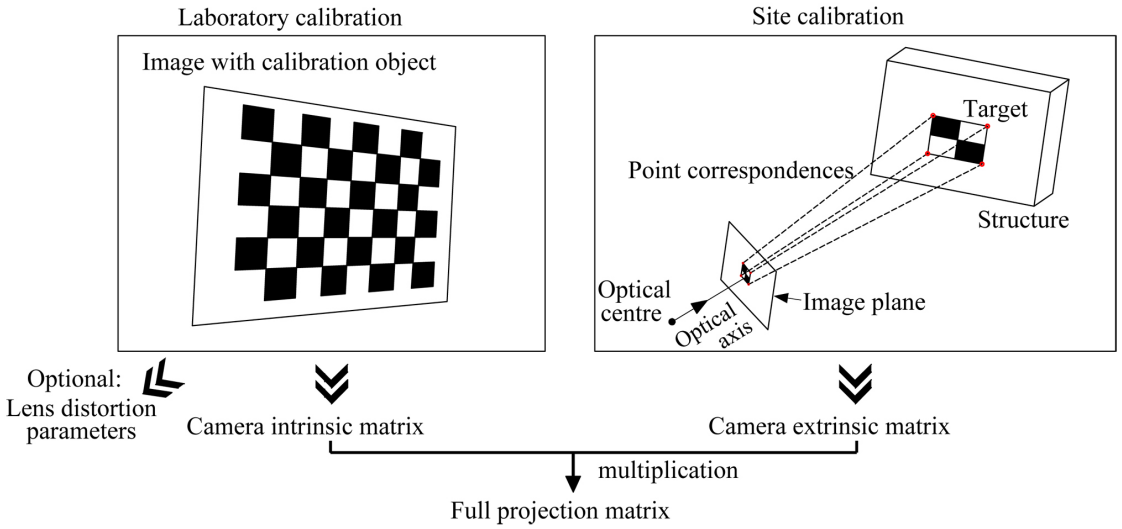


Fig. 2.4 Calibration steps for estimation of full projection matrix.

particularly in the corner regions of the frame as shown in Fig. 2.5(a). The lens distortion parameters could also be estimated in laboratory calibration and applied to correct the image with the rectified one in Fig. 2.5(b). For cameras equipped with lenses producing no apparent lens distortion, the distortion correction step is not necessary. Naturally for the monitoring measurements, it is preferable to locate the target region in the central area of the field of view (Ribeiro et al., 2014) which suffers less lens distortion, as shown in Fig. 2.5(a). In the second step, the camera extrinsic matrix representing the camera position and orientation is estimated on site through point correspondences, i.e. 3D structural coordinates of control points and 2D image coordinates of their projections in an image. Given at least four sets of point correspondences, least-squares optimisation is used to find the best option of camera extrinsic matrix that minimises the total re-projection error between the detected image points and the calculated image projection points.

The calibration algorithms are available in the Vision System Toolbox of MATLAB and the open-source library OpenCV.

Application review Camera calibration for full projection matrix estimation is commonly used to measure 3D structural displacement, with a few examples illustrating the method: The procedures of laboratory and site camera calibration are described by Kim et al. (2006) in an application to structural displacement monitoring in a three-span bridge under truck-induced vibration. The viability of motion capture systems for the laboratory vibration measurement was verified (Park et al., 2015) using a T-shaped calibration wand for the estimation of camera extrinsic parameters. In the case of a long span bridge, Martins et al. (2015) applied the calibration method to measure the 3D structural displacement at mid-span with the assistance of a set of four active targets. The estimated camera parameters can be refined when multiple cameras with overlapped views are

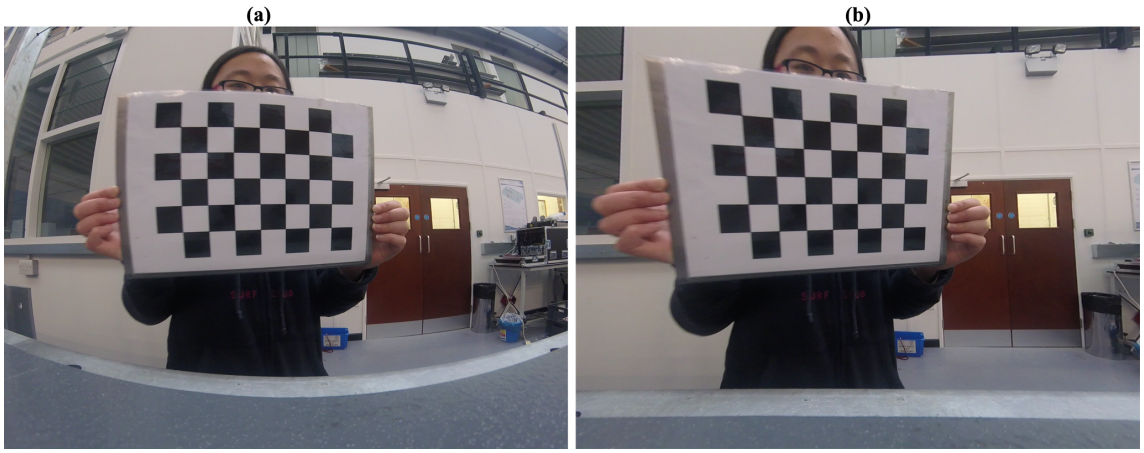


Fig. 2.5 Images of chessboard taken by GoPro Hero 4 Session camera: (a) raw image; and (b) image after distortion correction.

involved. For example, the methodology described by Chang and Ji (2007) is based on the epipolar geometry principle of stereoscopic vision where five points including structural point $\mathbf{P}_s$, projection points in two image planes $\mathbf{P}_I^1$ and $\mathbf{P}_I^2$, and two camera optical centres should all be coplanar, as shown in Fig. 2.6.

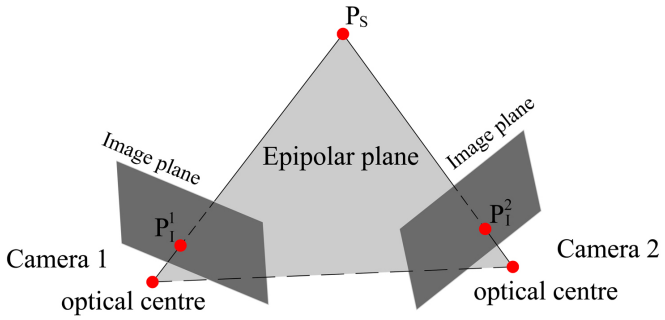


Fig. 2.6 Epipolar geometry principle of stereoscopic vision.

Remarks The full projection matrix is an accurate representation of the projection relationship and is thus applicable to any configuration of cameras on site. The lens distortion problems common for consumer-grade cameras do not prevent their use for such measurements, since corrections are readily made for distortion using laboratory camera calibration.

Camera calibration on site requires position information for some structural points. In existing studies this has been mainly acquired through the installation of artificial targets. Including artificial targets in laboratory tests is easy e.g. attaching a planar chessboard target (Chang and Ji, 2007; Kim et al., 2006) or placing a planar T-shaped wand in the field of view (Oh et al., 2015; Park et al., 2015) while the installation efforts in field tests (Martins et al., 2015) are much greater. The existing examples of two field applications are summarised in Table 2.3, indicating the feasibility of this method for both short-range and long-range monitoring tests.

Table 2.3 Summary of two field applications in literature using the full projection matrix as projection transformation

References	Chang and Xiao (2010b)	Martins et al. (2014, 2015)
Focal length	36.4 mm	600 mm (composed by a 300 mm telephoto lens and a 2x tele-converter)
Camera-to-target distance	5.2 m	500 m
Artificial targets installed	A planar 3x3 chessboard	A 3D target set combined by distributed four LED targets with the whole dimensions of 250 mm, 350 mm and 250 mm along the three axes
Observed maximum displacement	6 mm	1.82 m
Measurement evaluation	Not commented about vertical measurement; Measurement noise along the two other directions with the standard deviations at 0.76 mm and 1.09 mm, respectively.	Uncertainty at 15 mm to 20 mm in the vertical and transverse directions.

2.3.1.2 Planar homography

Principle For the case where the target moves within a plane contained in the 3D structural system (e.g. the XY plane), the projection relationship could be simplified to a planar homography between a 2D structural plane ($\mathbf{X}_p = [X, Y, 1]^T$) and a 2D image plane ($\mathbf{u} = [u, v, 1]^T$),

$$\alpha\{\mathbf{u}\} = [\mathbf{P}]_{3\times3}\{\mathbf{X}_p\} \tag{2.2}$$

where $[\mathbf{P}]_{3\times3}$ is the planar homography matrix and α is an arbitrary coefficient.

The reconstructed results using planar homography are usually the 2D structural displacement of targets. The calibration process requires at least four sets of 2D-to-2D point correspondences (Hartley and Zisserman, 2003), similar to the estimation process on site in full projection method.

Application review The planar homography considers the geometric distortion in the projection process and thus has no constraint on camera positioning (Lee et al., 2015). The 2D direct linear transform is effective for the planar homography estimation (Wu et al., 2014), for example the method was applied to monitor the oscillation of a laboratory steel frame with a dense array of markers glued to the

surface (Wu et al., 2014) and the mid-span deformation of a long-span bridge with an attached planar artificial target (Xu et al., 2016).

Remarks Planar homography applies no constraint on camera positioning and can be used to recover the target 2D structural displacements. In its application it is usual that the geometric information needed for calibration is provided by attaching artificial planar targets with known dimensions.

This calibration method is based on the assumption that the target moves within a structural plane with negligible motion along the third axis. Any motion not contained within this plane will lead to measurement error unless the motion is purely perpendicular to the camera optical axis.

2.3.1.3 Scale factor

Principle Scale factor is the simplest projection transformation and assumes an equal depth-of-field for all projected points or a camera configuration where the optical axis is perpendicular to one structural plane (Wu et al., 2014). With this assumption, the mapping process converts to a 1D-1D projection indicated in Fig. 2.7. The scale factor from the structural displacement to the image motion could be determined by one-dimensional correspondence or the camera-to-target distance,

$$\mathbf{SF} = \frac{\|\mathbf{P}_I\mathbf{Q}_I\|}{\|\mathbf{P}_S\mathbf{Q}_S\|} \tag{2.3}$$

or

$$\mathbf{SF} = \frac{f_{pix}}{D} \tag{2.4}$$

where $\|\mathbf{P}_I\mathbf{Q}_I\|$ and $\|\mathbf{P}_S\mathbf{Q}_S\|$ are the known physical dimension on the structural surface and the corresponding pixel length of the projection in image; f_{pix} is the camera lens focal length in terms of pixel units; and D denotes the distance from the camera optical centre to the structural surface plane.

For the system combining a camera with a total station, a projection coefficient called angular resolution (Charalampous et al., 2015; Ehrhart and Lienhart, 2015b) is used to perform the transformation which represents the angle value (α in Fig. 2.7) from the camera optical axis to a projection line with the projection length ($\|\mathbf{O}_I\mathbf{P}_I\|$) of one pixel. In principle, this projection transformation is similar to the scale factor estimated by camera-to-target distance in Equation (4) where the distance D is measured directly by electronic distance measurement (EDM) instrument and the focal length f_{pix} is related to the angular resolution θ by

$$\theta \approx \tan\theta = \frac{1}{f_{pix}} \tag{2.5}$$

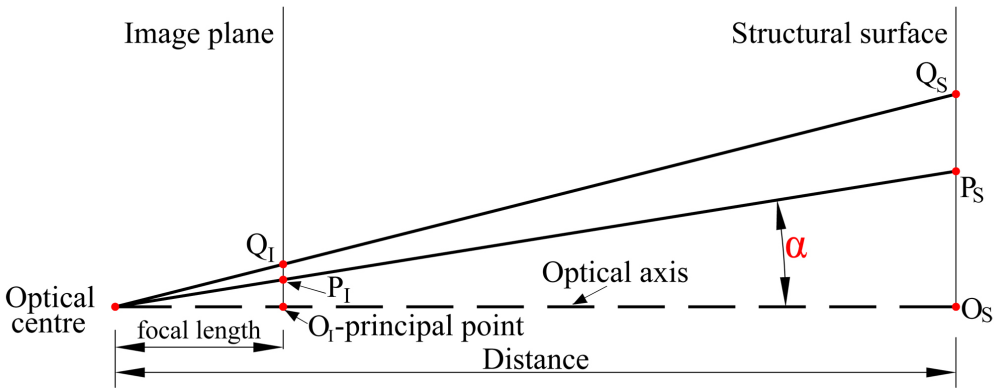


Fig. 2.7 Camera projection model when the optical axis of camera is perpendicular to the structural surface.

Application review Scale factor has been widely used to transform image motion to structural displacement with the features summarised in Table 2.4. Mostly the scale factor is determined via a known dimensions in an artificial target attached to structure (Choi et al., 2011; Feng et al., 2015a; Fukuda et al., 2010; Kim and Kim, 2011; Lee et al., 2006; Lee and Shinozuka, 2006; Olaszek, 1999; Ribeiro et al., 2014; Sładek et al., 2013; Stephen et al., 1993; Wahbeh et al., 2003; Ye et al., 2013; Yoon et al., 2016) while the method using the camera-to-target distance (Khuc and Catbas, 2017a) is less popular. For 2D structural displacement measurement, the scale factors for two axes within the target plane are calibrated separately according to dimension correspondences (Choi et al., 2011; Fukuda et al., 2010; Lee et al., 2006; Lee and Shinozuka, 2006). Error analysis indicates that the scale factor by dimension correspondence is less sensitive to the tilt of camera optical axis (Feng et al., 2015a). However, the scale factor using the camera-to-target distance has no dependence on artificial targets and thus is an easier way to realise completely non-contact monitoring (Khuc and Catbas, 2017a).

Table 2.4 Features of two calibration methods for scale factor

Scale factor	By camera-to-target distance	By dimension correspondences
Target dependence	Target free	Artificial targets always required
Camera positioning constraint	Very sensitive to the tilt of camera optical axis	Less sensitive to the tilt of camera optical axis
Applications	Mostly used in the short-range measurement; The long-range measurement feasible for the vision-based systems assisted the total station.	Widely used in both the short-range and the long-range measurement

Remarks Scale factor is the simplest projection transformation, particularly when no artificial target is used (Khuc and Catbas, 2017a; Yoon et al., 2016) and works when the camera optical axis is perpendicular to the structural surface. Camera positioning is less critical (Feng et al., 2015a) when known structural dimensions are used for calibration. However, when applying the scale factor derived from the camera-to-target distance, the tilt angle of the camera optical axis is suggested to be less than 10° through laboratory validation tests in short distance ($\leqslant$ 3.7 m) (Choi et al., 2016). Care must be taken that different scale factors are applied to different axes to measure the 2D displacement. This simple method can also be used with cameras having apparent lens distortion, since the lens distortion correction method previously described can be used (Sładek et al., 2013; Yoon et al., 2016).

2.3.2 Target tracking

Target tracking is the key part of a video processing package. In this study, target tracking techniques are categorised into four types based on target characteristics shown in Table 2.5, partly referring to Ehrhart and Lienhart (2015a).

Table 2.5 Categories of target tracking methods

Tracking methods		Regions or points tracked for matchings
(1)	Template matching	Rectangular subset of the frame as the target region
(2)	Feature point matching	Sparse 'special' points with salient features within the target region
(3)	Optical flow estimation	Every pixel location within the target region
(4)	Shape-based tracking	Line-type, circular-shaped or custom-made targets

2.3.2.1 Template matching

Principle Template matching is a classic technique for target tracking by searching in a new frame for an area most closely resembling a predefined template, following the procedures demonstrated in Fig. 2.8.

A rectangular region that is a subset in the reference frame is first selected as the template, and could be either an artificial target (Stephen et al., 1993) or a feature target on the structural surface (Feng et al., 2015a). A correlation criterion is required in order to evaluate the similarity level between the template and the new image subset. Robust criteria for matching are zero-mean normalised cross correlation (ZNCC) and zero-mean normalized sum of squared differences

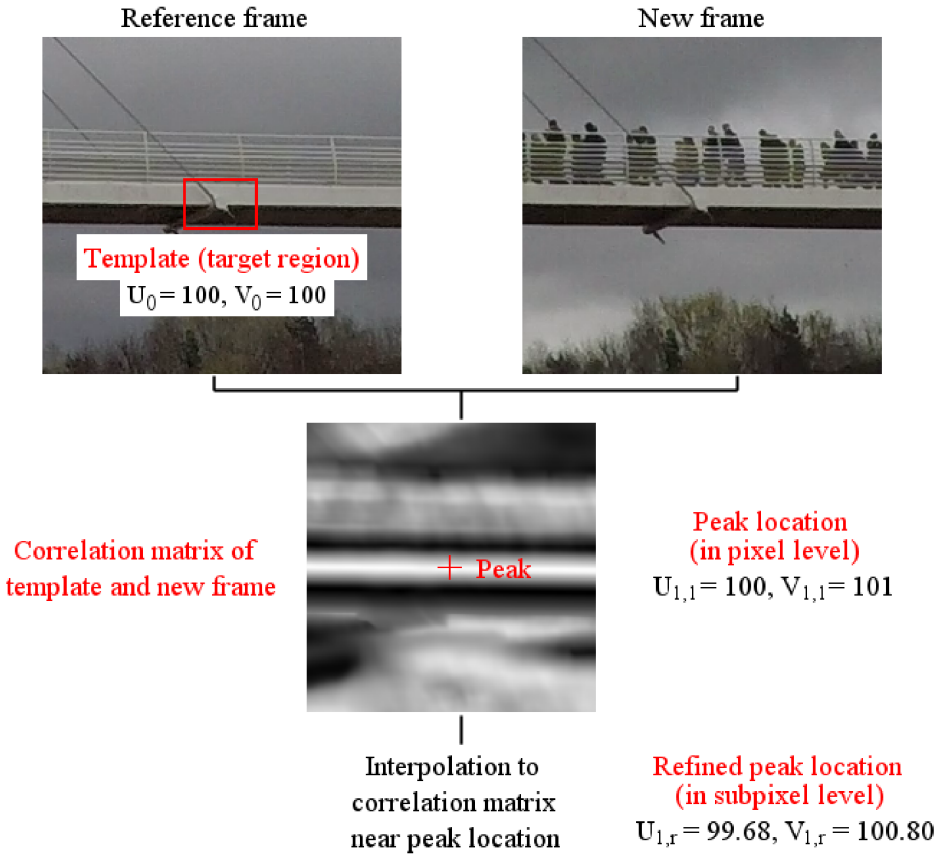


Fig. 2.8 Procedures of template matching method for target tracking: The horizontal and vertical coordinates of the target centre in the image plane are denoted as U and V, respectively; and the subscripts 0 and 1 represents the image coordinates in the reference and new frames, respectively.

(ZNSSD) which are insensitive to offset and linear scale in illumination (Pan et al., 2009) while another similarity criterion based on orientation code is also reported to be effective (Fukuda et al., 2013). The definition of the ZNCC criterion is provided as an example in Equation(2.6) while more details are given in (Pan et al., 2009).

$$\mathbf{C}_{ZNCC} = \sum_{i=-M}^{M} \sum_{j=-N}^{N} \frac{(f(x_i, y_i) - f_m)(g(x_i', y_j') - g_m))}{\Delta f \Delta g} \tag{2.6}$$

where $f(x_i, y_i)$ and $g(x_i', y_j')$ denote the image intensity values at the specified pixel locations in the template region and the new frame; f_m and g_m denote the average intensity values in the template region and the new frame; and Δf and Δg denote the standard deviations of intensity values in the template region and the new frame.

The location in the correlation map reaching the highest similarity is taken as the new image location of the target. The default resolution is at pixel level, so interpolation schemes (Feng et al., 2015a) are used to refine the result to the subpixel level. The feasible interpolation methods include bi-cubic interpolation

(Choi et al., 2011), second-order polynomial interpolation (Sładek et al., 2013) in spatial domain and zero-padding interpolation in frequency domain (Feng et al., 2015a). If the selected target includes robust and identifiable features, Harris corner detection that identifies the edge intersection points through a score value related to the eigenvalues of image gradient matrix could be an alternative to refine the initial matching location (Chang and Ji, 2007).

Application review Template matching is an established method widely applied in structural monitoring from the earliest work on the Humber and Severn Bridges in 1990s (Macdonald et al., 1997; Stephen et al., 1993). Recent applications include displacement monitoring tests on a railway bridge (Feng et al., 2015a), a long-span bridge (Ye et al., 2013) and a high-rise building (Liao et al., 2010).

Digital image correlation (DIC) is an extension of template matching mostly used in experimental mechanics (Pan et al., 2009; Schreier et al., 2009), with the difference that DIC considers the shape distortion under large deformation (Baker and Matthews, 2004) i.e. Lucas-Kanade template matching (Lucas and Kanade, 1981). As an example, a short-span railway bridge monitoring exercise (Guo and Zhu, 2016) used normalised correlation-based matching and Lucas-Kanade matching and indicated high similarity in both time and frequency domain.

Remarks Template matching is easy to use without user intervention apart from the initial selection of the template region. It does not have any special requirement for target patterns and has been validated to work well to track artificial planar targets with specific patterns (Chang and Ji, 2007; Macdonald et al., 1997; Stephen et al., 1993), LED lamp targets (Ye et al., 2013) and feature targets on structural surfaces (Feng et al., 2015a).

Template matching is not robust to changes in shading, illumination (Brownjohn et al., 2017; Guo and Zhu, 2016) and background conditions (Ehrhart and Lienhart, 2015c) in field, although sensitivity to lighting changes might be reduced using camera auto-exposure settings (Brownjohn et al., 2017). The method is also not appropriate for tracking slender structural components, since the rectangular subset image used as a template might include background pixels with inconsistent motion.

2.3.2.2 Feature point matching

Principle Instead of analysing all the locations within the target, feature point matching applies to sparse ‘special’ points within the target region, independently detecting these special points in two images and then finding point correspondences based on their local appearance. ‘Special’ points in an image, termed

'interest points' or 'keypoints' in computer vision, are those which are stable, distinctive and invariant to image transformation and illumination changes, such as building corners, connection bolts, or other patches with interesting shapes (Szeliski, 2011).

The procedures are indicated in Fig. 2.9. A popular keypoint detector in step (1) is the Harris corner detector (Harris and Stephens, 1988) which is widely used in structural monitoring applications (Chang and Ji, 2007; Khuc and Catbas, 2017a; Sładek et al., 2013; Yoon et al., 2016). Instead of using the pixel values directly for similarity comparison, keypoints are often extracted and described by a more complex representation (i.e. feature descriptor) according to the shape and appearance of a small window around the keypoint (Szeliski, 2011). The common descriptors and their matching criteria are indicated in Table 2.6. Float point based descriptors (e.g. scale-invariant feature transform (SIFT) (Lowe, 2004) and speeded up robust features (SURF) (Bay et al., 2008)) are represented by float vectors, commonly reflecting various local intensity gradients of a pattern around the keypoint. Binary string based descriptors (e.g. Binary robust independent elementary features (BRIEF) (Calonder et al., 2010), Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011) and Fast retina keypoint descriptor (FREAK) (Alahi et al., 2012)) are represented by binary vectors (with elements of 0 and 1) through pairwise comparisons of image intensities (i.e. whether the former is greater or less than the latter) over a special pattern around the keypoint. The matching criterion between two binary descriptors is usually the Hamming distance (Calonder et al., 2010) equal to the number of elements which differ between the two vectors.

To verify the matched keypoint correspondences in step (3), geometric alignment is often used based on whether the keypoints in the first image could fit with the keypoints in the second image after a specific geometric transformation. The widely used approaches for discarding outliers are RANdom SAmple Consensus (RANSAC) (Fischler and Bolles, 1981) and least median of squares (LMS) (Rousseeuw, 1984). The tracking output is the average motion of keypoints in an image that inherently has sub-pixel resolution and could be converted to the target location in the image.

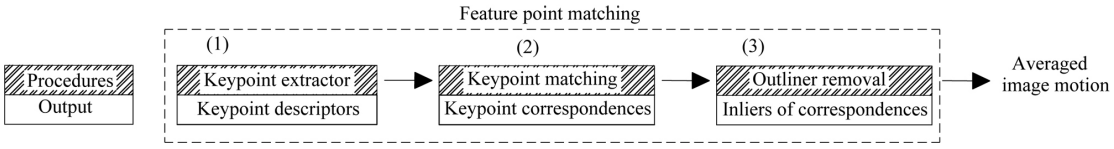


Fig. 2.9 Procedures of feature point matching for target tracking.

Application review Song et al. (2014) proposed a target tracking method based on circular Hough transform for marker detection and coherent point drift algorithm

Table 2.6 Categories of feature descriptors and corresponding matching criteria

Descriptor categories	Descriptor names	Matching criteria
Float point based	Scale-invariant feature transform (SIFT) (Lowe, 2004) Speeded up robust features (SURF) (Bay et al., 2008)	Euclidean distances in feature space (Szeliski, 2011)
Binary string based	Binary robust independent elementary features (BRIEF) (Calonder et al., 2010) Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011) Fast retina keypoint descriptor (FREAK) (Alahi et al., 2012)	Hamming distance (Calonder et al., 2010)

for marker matching and the method was applied for system identification of a steel cantilever beam in the laboratory. Field applications include Khuc and Catbas (2017a,b) who applied the FREAK and SIFT methods for deformation measurement in a stadium structure and a railway bridge and Ehrhart and Lienhart (2015a,c) who applied the ORB method for deformation measurement in a short-span footbridge.

Remarks Feature point matching is an efficient technique since it deals with sparse points instead of the whole region as in template matching. This technique uses local descriptors to represent keypoints instead of the raw image intensities and are less sensitive to illumination change, shape change and scale variation. However, feature point matching requires the target region to have rich textures for distinctiveness during the whole recording period. Also several parameters need to be adjusted manually according to users' experience or judgement, e.g. contrast threshold for feature detector and distance threshold in matching criteria. These parameter adjustments might depend on environmental changes, e.g. the threshold for outlier removal might vary with the illumination condition (Khuc and Catbas, 2017a).

Currently feature point matching technique has only been validated in several short-range measurement tests (Ehrhart and Lienhart, 2015a,c; Khuc and Catbas, 2017a,b). However, the feasibility for long-range monitoring in terms of stability over several hours and how to choose the best feature descriptors are open questions.

2.3.2.3 Optical flow estimation

Principle Instead of finding matching locations of a complete region or sparse keypoints, optical flow estimation detects motions or flows of all pixels within the

target region. Optical flow is the apparent velocity of movement in an image resulting from brightness pattern shift (Beauchemin and Barron, 1995). The calculation imposes two constraints, one temporal constraint on image properties (e.g. image intensity constancy for the same pattern over time) and one spatial constraint that models the flow properties in an image (e.g. coherent motion in adjacent pixels) (Sun et al., 2010). A function reflecting these two constraints is then defined and optimised to derive a dense estimation of velocity flow for each pixel. In structural monitoring applications, the output could be converted to image motion instead of velocity by replacing the temporal gradient of image properties in the function with the variation of image properties between two discrete frames. Outlier removal is used to retain only sensible image motions, and average image motion of the inlier pixels is converted to target location inherently having sub-pixel resolution.

Optical flow estimation is an established method with several variant techniques, such as 'differential', 'spatiotemporal energy' and 'phase-based'. In this section only two methods, the differential technique of Lucas and Kanade (LK) (Tomasi and Kanade, 1991) and the phase-based technique (Fleet and Jepson, 1990) are discussed.

LK optical flow estimation (Tomasi and Kanade, 1991) is based on brightness constancy assumption, i.e. projection of the same point has the same image intensity in every frame. Since corner points or keypoints are good features mathematically for the computation of optical flows, LK method is usually applied for sparse estimation instead of computation at every pixel. With keypoints detected in the reference frame usually using the Shi-Tomasi corner detector (Shi et al., 1994), LK algorithm is applied to compute the image motion of each keypoint in the new frame from spatial-temporal image brightness variations,

$$\begin{bmatrix} \sum_i I_{xi}^2 & \sum_i I_{xi} I_{yi} \\ \sum_i I_{xi} I_{yi} & \sum_i I_{yi}^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} -\sum_i I_{xi} I_{ti} \\ -\sum_i I_{yi} I_{ti} \end{bmatrix} \tag{2.7}$$

where dx and dy denote the optical flows in the horizontal and vertical directions of the image plane; I_x , I_y and I_t represent the spatial and temporal gradients of image intensities; and i denotes the ith pixel location in a square patch (e.g. 3×3) around a feature point (x, y). The image motion is then estimated after discarding false motion estimates according to RANSAC or LMS, as with feature point matching.

Phase-based optical flow estimation is based on local phase constancy assumption. The method first proposed by Fleet and Jepson in 1990 (Fleet and Jepson, 1990), is receiving new attention together with the motion magnification technique (Wadhwa et al., 2013) which visualises motions in image sequences that are not visible to the naked eye. The mathematical details of phase-based

optical flow estimation are explained in (Chen et al., 2015b) and the algorithm is briefly summarised here.

The Fourier shift theorem indicates that a delay of a signal in the time domain corresponds to a linear phase variation in the frequency domain. Similarly, the image motion in spatial domain is also reflected in phase changes in spatial frequency domain. The phase here is the local phase (Bülow, 1999) corresponding to a specific spatial location instead of the whole image, usually derived by employing a quadrature pair of filters consisting of an even real part and an odd imaginary part (Boukerroui et al., 2004) i.e. Gabor filters (Gautama and Van Hulle, 2002) and Gaussian derivative filters (Chen et al., 2015b) (demonstrated in Fig. 2.10). The image motion at each pixel is then estimated from the spatial-temporal variations of the local phase for the filtered image.

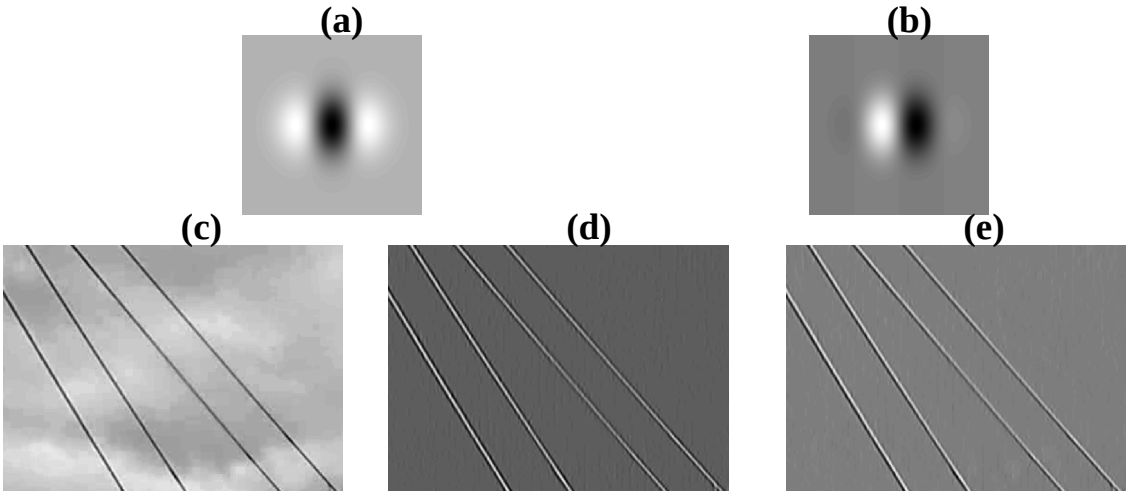


Fig. 2.10 Image after filtering by a quadrature pair of Gaussian derivative filters in the image width direction: (a) the real part of Gaussian derivative filters; (b) the imaginary part of Gaussian derivative filters; (c) the raw image of footbridge stay cables; (d) the real part of filtered image data; and (e) the imaginary part of filtered image data.

Application review LK optical flow estimation was applied in a laboratory test of a multi-storey metal tower (Yoon et al., 2016) for system identification, and for field application in deformation measurement in a footbridge (Ehrhart and Lienhart, 2015a) and bridge stay-cable vibration measurement (Caetano et al., 2011; Ji and Chang, 2008a).

Implementations of phase-based optical flow estimation were mostly focused on system identification, i.e. extracting modal frequencies and mode shapes in laboratory tests (Chen et al., 2015b; Yang et al., 2017) and identifying modal frequencies of high-rise tower buildings (Chen et al., 2017).

Remarks Optical flow estimation enables tracking of features on a structural surface without the requirement for artificial targets. It provides fast and accurate results in controlled environmental conditions.

Like feature point matching, optical flow estimation prefers target patterns with distinct and robust features over the whole test period. Edges are not suitable for tracking due to the 'aperture problem' i.e. only the motion component perpendicular to the local edge direction could be detected instead of the true motion of the edge. If the structural motion along edges is one dimensional translation with known direction e.g. bridge stay cable vibration (Ji and Chang, 2008a), optical flow estimation is viable.

Phase-based optical flow estimation is mostly applied for system identification in the laboratory but is harder to use in field conditions due to high signal noise (Chen et al., 2017). Measurement of image motion is sensitive to the choice of pixel location (Diamond et al., 2017), while a selection strategy to ensure satisfactory measurement has not yet been clearly reported. Changes of lighting and background conditions might lead to apparent measurement error (Chen et al., 2017).

2.3.2.4 Shape-based tracking

Other than general techniques, there are also some target tracking methods that depend on the special shapes of target patterns which could appear in custom-made artificial targets or structural components (e.g. line-like cables). Table 2.7 provides a summary of target patterns commonly used. With lack of generality, these methods have limitations for application.

2.3.2.5 Summary of target tracking performance

In terms of target tracking, the nominal algorithm resolution can be better than 0.01 pixel while the reported accuracy in practice varies from 0.5 to 0.01 pixel (Bing et al., 2006). The real-time processing was realised in (Feng et al., 2015a; Guo and Zhu, 2016; Xu et al., 2016) using the template matching method, in (Caetano et al., 2007, 2011) using the optical flow estimation method and in (Ehrhart and Lienhart, 2015b; Lee et al., 2006; Olaszek, 1999; Ye et al., 2013) using the shape-based tracking approaches. Although not reported in the existing applications, the feature point matching approach is capable of being used for real-time application (Rublee et al., 2011). Among the four tracking methods, template matching requires the least user intervention apart from the initial selection of template region while in the other three methods, some threshold values that might be environmentally dependent are required as user inputs.

Ehrhart and Lienhart (2015c) evaluated the performance of three techniques (optical flow, template matching and feature point matching) by tracking structural features of a footbridge and reported that feature point matching is robust to the changes of background condition (i.e. snowfall) whereas drift over time was observed in the measurement by the two other methods. Busca et al. (2014) evaluated three techniques (template matching, edge detection and digital image correlation) on a steel truss railway bridge, concluding that the three techniques provide similar tracking performance while tracking accuracy is slightly poorer for natural targets. Khaloo and Lattanzi (2017) investigated four optical flow estimation methods for dense displacement measurement. The study indicated that classic+NL method (i.e. introducing a weighted non-local term into the classical optical flow formulations (Sun et al., 2010)) provided the most consistent and accurate measurement. However, the coarse-to-fine schemes (i.e. building image pyramids for each frame and computing optical flows on each layer of pyramids to get rid of the small motion constraint) are necessary for Lucas–Kanade and Horn–Schunck methods to deal with large displacement.

2.3.3 Structural displacement calculation

Structural displacement could be easily derived from the change of structural coordinates given the image location of a target (output of target tracking) and the projection transformation relationship (output of camera calibration). In this case, the projection transformation is a fixed value or matrix without any update during the test.

Another less common method to derive structural displacement is based on the variation of real-time camera extrinsic matrix. The camera extrinsic matrix represents the camera pose i.e. position and orientation relative to the structural system. Since the camera is physically fixed during the recording, variation of camera extrinsic matrix is related to the change of target pose (position and orientation) and could be used to estimate the target motions in six degrees of freedom (6DOF).

2.3.3.1 Offline projection transformation

Principle For single camera applications using scale factor or planar homography, the 2D structural coordinate/displacement is derived uniquely through transforming the target location/motion in an image to that in the structure via a projection transformation value or matrix.

When two or more cameras with overlapped views are used to monitor the same target, 3D structural displacement can be extracted based on triangulation method (Hartley and Zisserman, 2003).

Application review Applications of scale factor and planar homography for 2D structural displacement measurement have been reviewed in Section 2.3.1.

For stereoscopic view or multiple cameras, the triangulation method was used in (Chang and Ji, 2007; Jáuregui et al., 2003; Santos et al., 2009) for 3D structural displacement measurement. A multi-camera arrangement provides more reliable results than a single view but the measurement quality has high dependency on the time-synchronisation of camera recordings.

2.3.3.2 Online pose estimation

Principle For single camera applications, using a fixed projection transformation relationship only supports recovery of 2D structural displacement. Some researchers tried to extract more information about target motion (up to 6DOF) using a single camera by updating the real-time target pose in the structural system.

Estimation of camera extrinsic matrix is performed for every frame and the 3D translational and rotational target motions are extracted from the changes of camera extrinsic matrix compared to the initial frame. The calibration process requires at least four non-collinear points with known dimensions or spacing in structure that should have consistent motion.

Application review Greenbaum et al. (2016) applied the online pose estimation method for the laboratory 3D motion measurement of an oscillating rigid object with a few targets of known positions distributed on its surface. In field applications, Chang and Xiao (2010b) used a planar target with square grid patterns attached to a bridge surface for the measurement of 6DOF structural displacement while Martins et al. (2015) tracked four non-coplanar LED targets together to reconstruct the 3D structural motion in a long span bridge.

Remarks The greatest advantage of the method is the capacity to extract 6DOF structural motions from single camera, but it has a high requirement on the nature of tracked targets which should consist of at least four non-collinear points with precisely known geometry. The target or a set of target points should have rigid motions and be visible during the whole recording period e.g. artificial planar targets with salient corner points (Chang and Xiao, 2010b), distributed target points on structural surface (Greenbaum et al., 2016) or a set of LED targets (Martins et al., 2015).

This technique cannot measure translation along the camera optical axis (Chang and Xiao, 2010b), thus the camera should be configured to avoid facing any motion direction of interest. The measurement accuracy of this method might be poorer than offline projection transformation method. In a footbridge monitoring

test by Chang and Xiao (2010b), using a 36.4 mm focal length camera placed about 5.2 m from mid-span generated measurement noise with standard deviations of 0.76 mm and 1.09 mm in two horizontal directions.

This was much larger than would be achieved by offline projection transformation method in a similar test (Brownjohn et al., 2016b) (tracking 0.2 mm bridge vertical displacement with the 85 mm focal length and 27 m camera-to-target distance). Therefore this method is not recommended for field applications unless the target size is not negligible compared to the camera-to-target distance (Busca et al., 2014).

2.4 Field applications and challenges

This section summarises the existing field applications of vision-based systems in two active fields, bridge deformation measurement and cable vibration measurement. A discussion about measurement challenges in field applications is also presented.

2.4.1 Application examples

Video acquisition devices are now expanded to include smartphone cameras, with numerous applications including vibration measurement of a laboratory multi-floor tower structure (Yoon et al., 2016) and cable vibration measurement of a cable-stayed footbridge (Zhao et al., 2017). In these two applications, smartphones are only used as the data acquisition system with the recorded video files post-processing for data extraction. Smartphone applications for real-time video acquisition and processing are also viable (Min et al., 2016) through experimental validations.

The existing applications of vision-based systems in field tests involve the deformation measurement of a wide range of structural types including: high-rise buildings (Chen et al., 2017; Liao et al., 2010), bridges (Brownjohn et al., 2017; Chang and Xiao, 2010b; Ehrhart and Lienhart, 2015a,b,c; Feng et al., 2015a; Feng and Feng, 2015, 2017; Feng et al., 2015b; Fukuda et al., 2010; Guo and Zhu, 2016; Khuc and Catbas, 2017a,b; Kim et al., 2006; Lee et al., 2006; Lee and Shinozuka, 2006; Macdonald et al., 1997; Martins et al., 2014, 2015; Ojio et al., 2016; Olaszek, 1999; Ribeiro et al., 2014; Santos et al., 2009; Stephen et al., 1993; Wahbeh et al., 2003; Xu et al., 2016; Ye et al., 2013; Yoneyama and Ueda, 2012) and stadium structures (Chang and Ji, 2007; Feng et al., 2017). Work in the two most active fields, i.e. bridge deformation measurement and cable vibration measurement are summarised in Table 2.9 and Table 2.8, respectively.

The viability of vision-based systems for bridge displacement measurement has been verified through comparison with traditional displacement sensors, e.g. LVDT (Feng et al., 2015b; Fukuda et al., 2010; Ribeiro et al., 2014), laser sensors (Fukuda et al., 2010) and potentiometers (Kim et al., 2006) for short-span bridge and GPS (Brownjohn et al., 2017; Xu et al., 2016; Ye et al., 2013) for long-span bridges. The displacement data could be interpreted for system identification (Chang and Xiao, 2010b; Feng et al., 2015a; Fukuda et al., 2010; Guo and Zhu, 2016; Khuc and Catbas, 2017b; Lee and Shinozuka, 2006; Wahbeh et al., 2003), evaluation of load carrying capacity (Lee et al., 2006), model calibration (Feng and Feng, 2015) and estimation of vehicle weights (Ojio et al., 2016). Artificial targets are commonly used in existing applications to assist camera calibration, whereas recent investigations (Ehrhart and Lienhart, 2015b; Feng and Feng, 2017; Feng et al., 2015b; Guo and Zhu, 2016; Khuc and Catbas, 2017b) overcome the dependence on artificial targets and realise completely non-contact sensing based on a simplified projection transformation i.e. scale factor.

Another promising application of vision-based systems is to estimate cable tension forces based on vibration measurement. Measurement accuracy was verified through comparison work with traditional sensors e.g. accelerometers (Ji and Chang, 2008a; Kim et al., 2013; Kim and Kim, 2013), velocimeters (Chen et al., 2015a) and load cells (Feng et al., 2017). Vision-based systems require no access to cables (Brownjohn et al., 2017; Caetano et al., 2007, 2011; Feng et al., 2017; Ji and Chang, 2008a,b; Kim et al., 2013; Kim and Kim, 2013) and are capable of measuring the vibrations of multiple cables using a single camera (Caetano et al., 2007, 2011; Feng et al., 2017; Kim et al., 2013) that is comparable to an array of accelerometers.

2.4.2 Measurement accuracy and challenges

Measurement accuracy of vision-based systems depends on several parameters, e.g. camera-to-target distance, target pattern features, lighting conditions, camera mounting stability and video processing methods. Khuc and Catbas (2017a) investigated the measurement accuracy of a vision-based system in a laboratory and suggested an accuracy of 0.04 mm in a short-range distance (< 14 m). Martins et al. (2015) demonstrated the uncertainty evaluation of displacement measurement by a vision-based system on a long-span bridge monitoring test and illustrated a standard measurement accuracy of 1.7 mm in the vertical direction. The high noise range might limit the field application of vision-based systems for system identification on civil structures although high frame rate is taken for vision-based systems.

The achievable accuracy in field tests might be much poorer than that of controlled conditions. The authors investigated the field challenges through a

series of monitoring tests in two short-span and two long-span bridges which have been reported in Brownjohn et al. (2017). A summary of the main findings from the tests and the literature is presented here.

- Camera and support motion induced by wind (Ribeiro et al., 2014) might lead to apparent measurement error. Except for improving camera mounting configurations (Brownjohn et al., 2017), a common correction method is to additionally track the 'nominal' motion of a fixed target e.g. bridge towers or adjacent buildings. Recent work (Khaloo and Lattanzi, 2017) indicates another promising approach for camera motion compensation through removing the averaged motion of background pixels based on dense optical flow estimation.

- Variation in lighting and background conditions is one of the critical challenges during field tests. The influence of lighting variations might be reduced by enabling camera auto-exposure settings (Brownjohn et al., 2017). Correlation-based template matching is not robust to this effect apart from testing during overcast weather, whilst the feature point matching method was reported to be less sensitive (Ehrhart and Lienhart, 2015c).

- Atmospheric refraction and turbulence of optical light propagating through the air are common error sources for any optical-based instrument, especially for long-range measurements. Refraction deviation could be minimal for short-term displacement measurement while the air turbulence movement has a larger influence (Olaszek, 1999). Quantification of the induced error based on mathematical models is demonstrated in a vision-based measurement test of a long-span bridge (Lages Martins et al., 2014).

- Observations from short-term tests (with duration less than twelve hours) do not find an apparent influence of temperature variations on measurement accuracy, while this effect is necessary to consider for long-term tests e.g. with duration a few months or more. A time-frequency approach indicates the potential for error compensation based on investigation of the correlation models linking measurement errors and temperatures (Zhou et al., 2017).

2.5 Summary and prospects

As evidenced from the review, vision-based systems are promising tools for structural displacement measurement having advantages in cost, installation efforts and measurement capacities of frequency range and spatial resolution. Although the potential in field applications has been validated in many articles, there are a few aspects still to mature.

- Robust target tracking methods. Template matching and optical flow estimation are established methods widely used in short-range and long-range measurement tests, but they are not robust to lighting and background changes. Feature point matching is a relatively new and promising tracking method, but investigations regarding several aspects e.g. selection strategy of proper threshold parameters, sensitivity on environmental effects and field viability for long-range measurement are rare, and need to be expanded. It is still an open question about the most robust tracking method for vision-based systems to deal with changes in lighting conditions during field tests.

- Completely non-contact measurement. Artificial targets are commonly included to assist the camera calibration process, but dependence on artificial targets is eliminated in a few field applications (Feng and Feng, 2017; Khuc and Catbas, 2017a,b; Yoon et al., 2016). These studies were based on a simplified projection transformation i.e. scale factor that is not a general approach and imposes constraints on camera positioning. To develop a non-contact vision-based system for the general case, requiring control points with known locations is the main obstacle which could possibly be resolved via the assistance of surveying instruments, such as total station.

- Distributed sensing of structural displacement. Vision-based systems have the capacity for simultaneous multi-point displacement measurement that is comparable or superior to an array of accelerometers for system identification. Currently, bridge applications primarily focus on the mid-span displacement measurement, while the potential of distributed sensing and system identification is not well investigated.

- Measurement uncertainty evaluation. Measurement accuracy and uncertainty are of great importance for a mature measurement system. Quantified descriptions about measurement accuracy haven been made in some references (e.g. (Ehrhart and Lienhart, 2015c; Feng et al., 2015a; Khuc and Catbas, 2017a; Lee and Shinozuka, 2006) through comparisons with reference measurements. However, the quality of vision-based measurements could be time-varying, environmentally dependent and differ significantly with various test configurations. The influential factors include the test configurations (e.g. the camera-to-target distance and the target features), the video processing methods used and the environmental conditions (e.g. the lighting conditions, the atmospheric refraction and turbulence). A systematic evaluation of vision-based measurement methodologies will require extensive experimental effort by the research community with publication of case studies contributing to evolving guidance for field applications.

Table 2.7 Examples of shape-based tracking

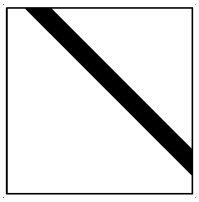
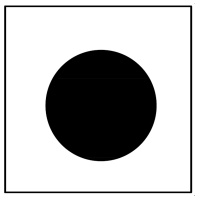
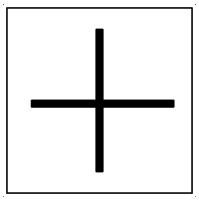
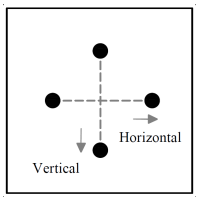
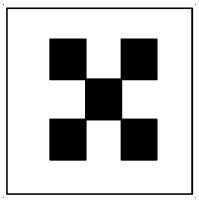
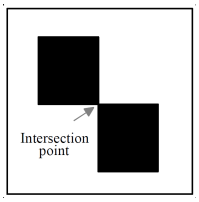
References	Target patterns	Determination of target location
Brownjohn et al. (2017); Ji and Chang (2008b)		Detecting the edges of line-shaped patterns and building image point correspondences among image sequences (Ji and Chang, 2008b) or computing the cable motion from the distance between two identified edges (Brownjohn et al., 2017)
Chen et al. (2015a); Park et al. (2014); Shan et al. (2015); Wahbeh et al. (2003)		Detecting the edges of circular-shaped patterns through brightness thresholding or edge detection and computing the centroid coordinates for the circle
Olaszek (1999)		Detecting the edges of cross-shaped patterns through image gradient and computing the arithmetic mean of edge coordinates as the target location
Fukuda et al. (2010); Ho et al. (2012); Lee et al. (2006); Lee and Shinozuka (2006); Yi et al. (2013a)		Detecting four spots through brightness thresholding and computing the motions along the specified horizontal and vertical directions
Chang and Xiao (2010b)		Detecting grid dots by Harris corner detector and applying the image coordinates of grid dots for the estimation of camera extrinsic matrix
Ribeiro et al. (2014)		Detecting the edges of squares through brightness thresholding and computing the coordinates of the intersection point

Table 2.8 Review of studies about cable vibration measurement using vision-based systems

References	Application structures	Target tracking methods	Data interpretation
Ji and Chang (2008a)	A footbridge	Optical flow estimation	Extracting modal frequencies
Ji and Chang (2008b)	A footbridge	Shape-based tracking	Extracting modal frequencies Identifying mode shapes
Caetano et al. (2011)	Guadiana Bridge, Portugal	Optical flow estimation	Extracting modal frequencies
Kim et al. (2013); Kim and Kim (2013)	Gwangan Bridge and a two-pylon cable-stayed bridge in Busan-Geoje Fixed Link, Korea	Correlation-based template matching	Extracting modal frequencies Estimating cable tension
Chen et al. (2015a)	Chi-Lu Bridge, Taiwan China	Shape-based tracking	Extracting modal frequencies Identifying the mode shape ratio of cables
Feng et al. (2017)	Hard Rock Stadium, USA	Correlation-based template matching	Extracting modal frequencies Estimating cable tension
Brownjohn et al. (2017)	A footbridge	Shape-based tracking	Extracting modal frequencies
Zhao et al. (2017)	A footbridge	Edge detection	Extracting modal frequencies Estimating cable tension

Table 2.9 Review of studies about bridge displacement measurement using vision-based systems

References	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
Macdonald et al. (1997); Stephen et al. (1993)	Humber Bridge & Second Severn Crossing, UK Scale factor	Correlation-based template matching	Planar target	2D displacement at mid-span		
Olaszek (1999)	A highway bridge & a railway viaduct	Scale factor	Shape-based tracking	Planar target	Vertical displacement at mid-span	
Wahbeh et al. (2003)	Vincent Thomas Bridge, USA	Scale factor	Shape-based tracking	LED targets	Vertical displacement at mid-span	Extracting modal frequencies
Fukuda et al. (2010); Lee et al. (2006); Lee and Shinozuka (2006)	Highway bridges	Scale factor	Shape-based tracking	Planar target	2D displacement at mid-span	Estimating load carrying capacity for load test Evaluating measurement by comparison with the reference sensors (LVDT & Laser)

Continued on next page

– continued from previous page

References	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
Kim et al. (2006)	A roadway bridge	Full projection matrix	Shape-based tracking	Planar target	2D displacement at mid-span	Evaluating measurement by comparison with the reference sensor (potentiometer)
Chang and Xiao (2010b)	A cable-stayed footbridge	Online pose estimation	Shape-based tracking	Planar target	3D displacement at mid-span	Extracting modal frequencies
Ye et al. (2013)	Tsing Ma Bridge, Hong Kong	Scale factor	Correlation-based template matching	LED targets	Vertical displacement at mid-span	Evaluating measurement by comparison with the reference sensor (GPS)
Guo and Zhu (2016)	A railway viaduct	Scale factor	Lucas-Kanade template matching	Natural features	Vertical displacement of sound barrier	Extracting modal frequencies
Ribeiro et al. (2014)	A railway bridge	Scale factor	Shape-based tracking	Planar target	Vertical displacement at mid-span	Evaluating measurement by comparison with reference sensor (LVDT)

Continued on next page

– continued from previous page

References	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
Feng et al. (2015a); Feng and Feng (2015, 2017); Feng et al. (2015b)	A footbridge, a highway bridge & a railway bridge	Scale factor	Correlation-based template matching	Both planar target and natural features	Vertical displacement at mid-span	Extracting modal frequencies FE model calibration
Martins et al. (2014, 2015)	P25A bridge, Portugal	Online pose estimation	Shape-based tracking	LED targets	3D displacement at mid-span	Evaluating measurement uncertainty
Ehrhart and Lienhart (2015a,b,c)	A footbridge	Scale factor	Correlation-based template matching Feature point matching Optical flow estimation	Both planar target and natural features	Vertical displacement at mid-span	Extracting modal frequencies Evaluating measurement accuracy and stability of three tracking methods

Continued on next page

– continued from previous page

References	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
Brownjohn et al. (2017); Xu et al. (2016)	Humber Bridge, UK	Planar homography	Correlation-based template matching	Planar target	2D displacement at mid-span	Extracting modal frequencies Evaluating measurement by comparison with the reference sensor (GPS)
Ojio et al. (2016)	A roadway bridge	Scale factor	Lukas–Kanade method	Natural features	Vertical displacement at mid-span	Estimating vehicle weights
Khuc and Catbas (2017b)	A railway bridge	Scale factor	Feature point matching	Natural features	Vertical displacement at mid-span	Extracting modal frequencies

Chapter 3

Evaluation of a commercial vision-based system

This chapter presents field tests using an expensive commercial vision-based system (i.e. Imetrum Dynamic Monitoring Station) for deformation measurement on two long-span bridges (i.e. the Humber Bridge and the Tamar Bridge). The purpose of these two tests is to illustrate the potential and to build confidence in vision-based monitoring. A discussion about Imetrum system is presented in the end with respect to both the hardware and software. Application experiences here will be the basis for proposing a custom-developed vision-based system in the next chapter.

3.1 Imetrum Dynamic Monitoring Station (DMS)

The Vision System originating from research at the University of Bristol (Macdonald et al., 1997; Stephen et al., 1993) led to the 'Video Gauge' software that was commercialised via the university spin-out Imetrum formed in 2003. The implementation of the Video Gauge in a hardware platform is called the Dynamic Monitoring Station (DMS) which includes one or more GigE high performance cameras.

A typical implementation of the DMS for bridge monitoring is shown in Fig. 3.1. High-resolution cameras equipped with long focal length lens are connected to the controller (computer) via Ethernet cables and a group of up to four cameras are available for time-synchronised recording and real-time video processing. Cameras used in this study have a resolution of 2048 $\times$ 1088 pixels and sensor size of 11.26 $\times$ 5.98 mm.

In the video processing software, target tracking algorithms used are correlation-based template matching and super resolution techniques (Potter and Setchell, 2014) which enable better than 1/100 pixel resolution at sample rates beyond 100 Hz in field applications. The tracking objects could be either artificial targets

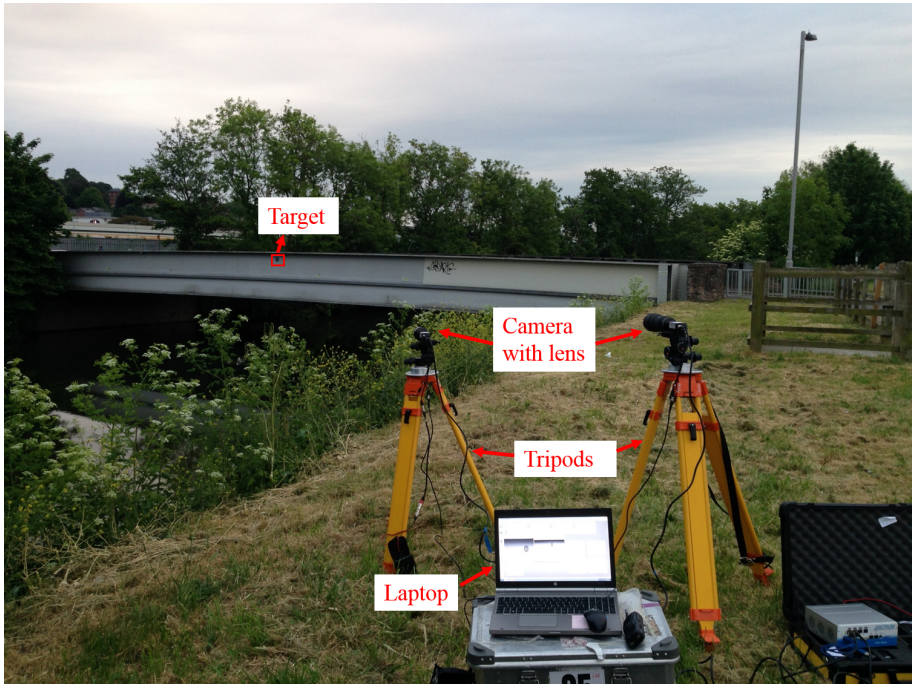


Fig. 3.1 Configuration of Imetrum DMS in a short-span bridge monitoring test.

or an existing feature (i.e. bolts or holes) on the bridge surface. The system supports either two-dimensional (2D) structural displacement measurement by single camera or three-dimensional (3D) structural displacement measurement by multiple cameras.

Before deploying on long-span bridges, the system has been validated on a short-span bridge through comparison with LVDT measurement. (Hester et al., 2017).

Section 3.2 and 3.3 report the field tests of operational deformation measurement on two long-span bridges, the Humber Bridge and the Tamar Bridge. Discussion about Imetrum DMS with respect to installation efforts, application ranges and measurement performance is given in section 3.4.

3.2 Field test on Humber Bridge

The Humber Bridge (Brownjohn et al., 2015), opened in 1981, has a main span of 1410 m and is known to experience mid-span deformations around 1 m in strong winds (Brownjohn et al., 2015).

A long term monitoring system has been operating at Humber Bridge since 2010 (Brownjohn et al., 2015) which includes a base station at the bridge tower, two GPS rover receivers (Leica GMX902) mounted on the main cables at mid-span and three QA750 accelerometers mounted inside the steel box girder at mid-span (two vertical and one lateral/horizontal) (Brownjohn et al., 2015). The sample rates

of GPS and QA accelerometers in the monitoring system are 1 Hz and 20 Hz, respectively.

A single day of field-testing using the DMS was performed to measure the lateral and vertical displacement at mid-span on a clear mid-summer day with low winds and normal traffic load. The DMS performance was evaluated in time and frequency domains by comparing with a GPS 'reference sensor'.

3.2.1 Test configuration

In the field test, the camera and controller along with battery power supply were located near the foundation of the North (Hessle) tower shown in Fig. 3.2(a), to the East of the pylon. A 300 mm f/2.8 lens was attached to both camera and tripod via a rigid double-support translation stage. A custom-made 1 m square steel frame holding an artificial target was mounted on the parapet at the mid-span 710 m from the lens as shown in Fig. 3.2(b). The pattern of the target is a set of concentric rings with a gradual blend from black to white at the edges.

The frequency range of interest containing the majority of significant vibration modes was less than 1 Hz, so the frame rate of the DMS system was chosen as 10 Hz. To save storage space, the image size of each frame was saved as 850 $\times$ 400 pixels although the default image size is 2048 $\times$ 1088 pixels.

Both the custom-made target and a natural feature target at mid-span were tracked. Fig. 3.2(c) shows a sample video frame. The red dashed boxes in the figure include the custom-made target and the natural target. The latter comprises ribs of the box deck on the bridge soffit and while it is judged to be close to the artificial target, it could be at a span-wise location differing by a few metres. To transform the image natural units (pixels) to the real world units (e.g. mm), a projection transformation reflecting the geometric relation between the 2D image plane and the 3D structural coordinate system i.e. planar homography is calibrated based on the known dimensions of the installed artificial target.

3.2.2 Measurement results

During the test, a D-SLR camera was adapted to video-record traffic on the bridge. Fig. 3.3(a) shows one captured frame from the video recorded when two heavy good vehicles (trucks) approached each other at mid-span from opposite directions. Fig. 3.3(b) shows the corresponding measurement using the DMS in the vertical direction, with vertical deflection at mid-span caused by the two vehicles reaching 221 mm. In general, the measurements from tracking each of the two targets agree well; the DMS demonstrates similar performance for tracking either target.

Displacement from tracking the custom-made target was used for comparison with the displacement data from one of the mid-span GPS receivers. Fig. 3.4(a)

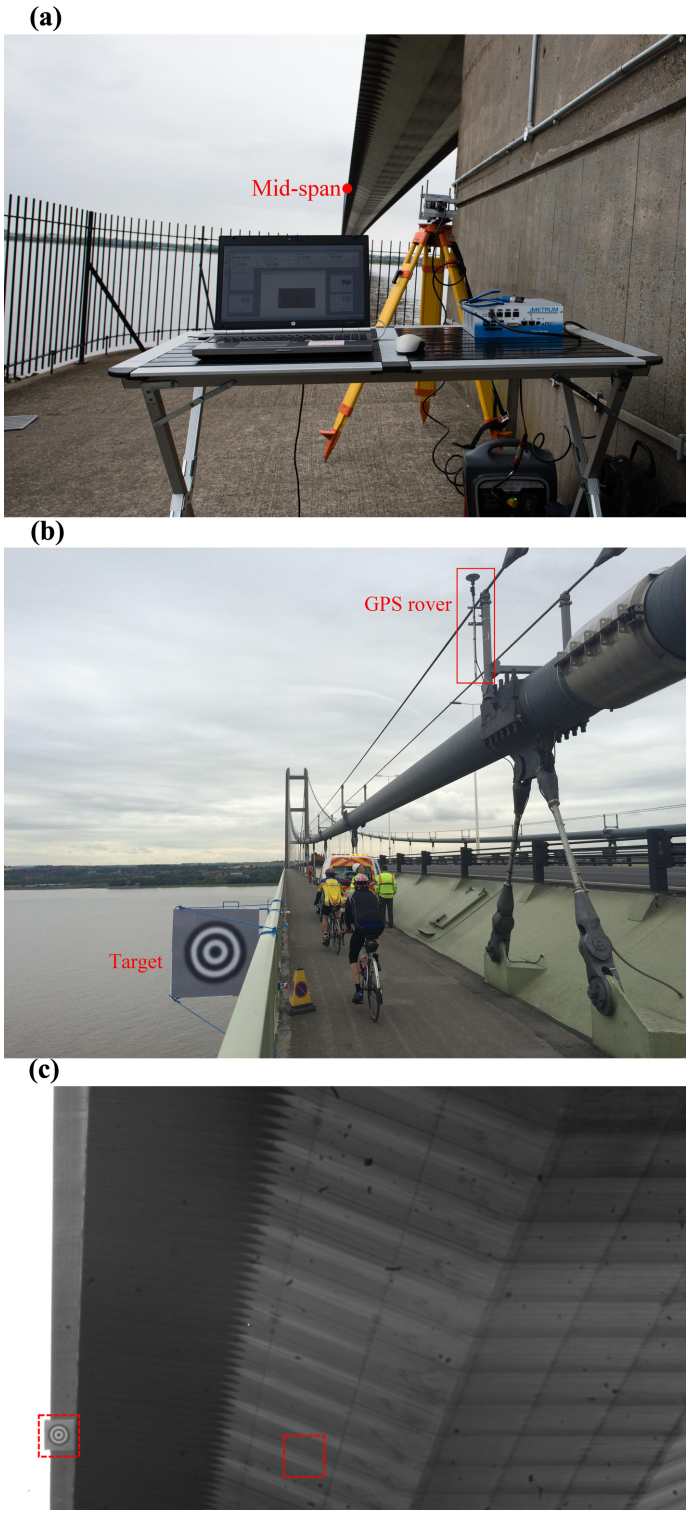


Fig. 3.2 Configuration of Imetrum DMS in the Humber Bridge test: (a) DMS near tower foundation; (b) custom-made target installed at mid-span; (c) captured frame from DMS video records with marked custom-made target and natural feature target

shows the vertical displacement measurement by DMS and the GPS receiver; the cross correlation coefficient of two signals is calculated to be 98.8%. Fig. 3.4(b) shows a zoom-in view of one-minute of data. Consistent with reported GPS standard accuracy of 35 mm in vertical direction (Nickitopoulou et al., 2006), the

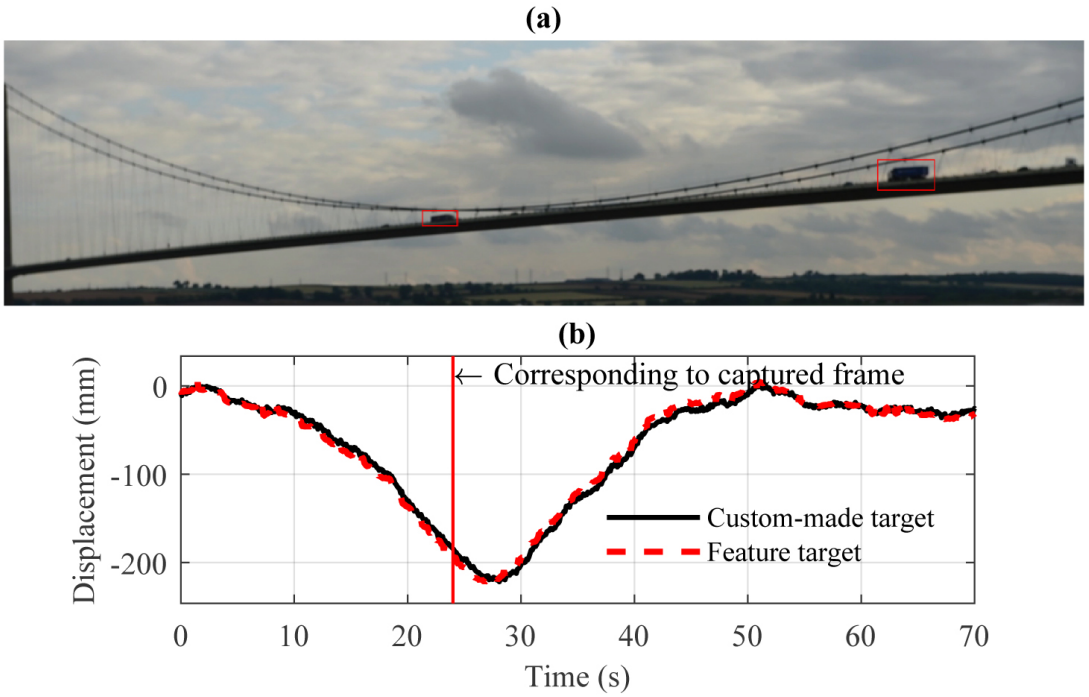


Fig. 3.3 DMS Measurement with two heavy vehicles approaching midspan from opposite directions: (a) one captured frame from video file by D-SLR camera; and (b) vertical displacement measured by DMS.

accuracy of the GPS observation at Humber was at the centimetre-level. For Imetrum DMS, it is hard to quantify the measurement accuracy on site since the true bridge motion is unknown. One approximate way to estimate the measurement accuracy is from target tracking accuracy using the scaling factor method,

$$\frac{I_{disp}}{S_{disp}} = \frac{f_{pixel}}{D} \tag{3.1}$$

where I_{disp} and S_{disp} are the target motions in the image plane (e.g. pixel) and structural system (e.g. mm); f_{pixel} is the focal length of camera lens in terms of pixel units corresponding the focal length in terms of mm scaled by the camera sensor resolution; and D denotes the distance between the camera optical centre and the target surface plane.

The nominal resolution of I_{disp} in target tracking can be better than 0.01 pixel while the reported accuracy varies from 0.5 to 0.01 pixel (Bing et al., 2006) which is related to target pattern (texture contrast) (Busca et al., 2014) and illumination condition. In this application, given the focal length of 300 mm, the camera sensor resolution at 0.0055 mm/pixel and the camera-to-target distance at approx. 710 m, the accuracy of 0.1 pixel (artificial target of high-contrast pattern) in image plane corresponds to an accuracy (or rather resolution) of 1.3 mm in the structural system.

The power spectral densities (PSD) of the DMS displacement signal, the GPS displacement signal and the acceleration data were obtained by Welch's method

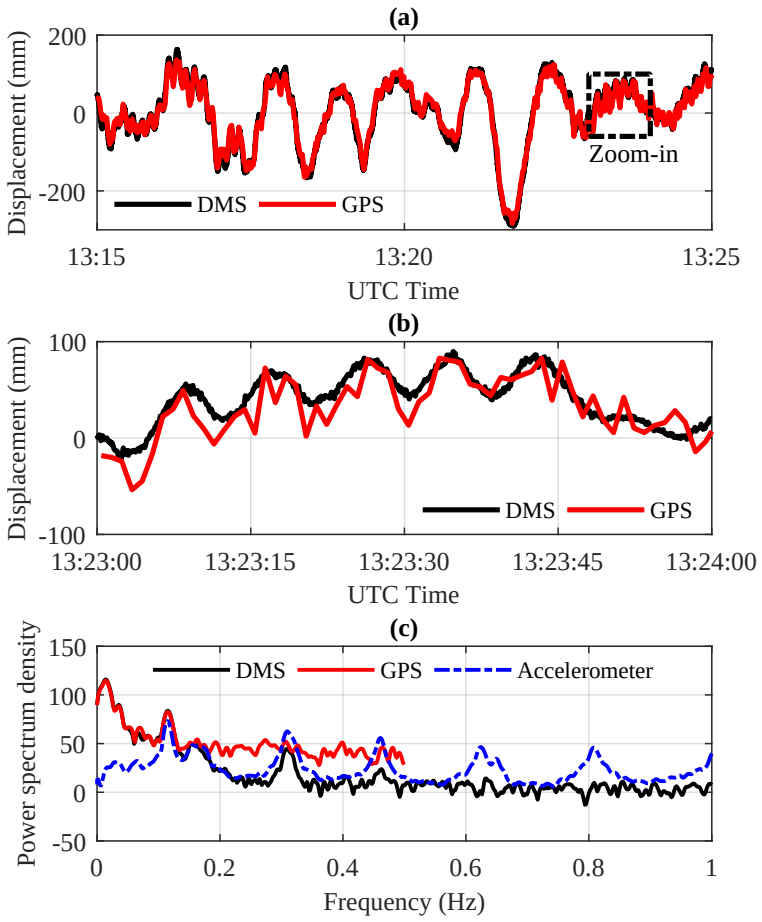


Fig. 3.4 Comparison of vertical displacement by DMS and GPS: (a) ten-minute signals of vertical displacement; (b) zoom of the area marked by rectangle in (a); and (c) power spectral density of displacement measurement signals.

shown in Fig. 3.4(c). From the previous studies on the bridge it is known that (vertical) modal frequencies exist at 0.117 Hz, 0.31 Hz and 0.46 Hz (Rahbari et al., 2015). The DMS captured the first and second vertical modal frequencies while the GPS captured only the first one. In theory, the DMS and GPS measurement sampled at 10 Hz and 1 Hz, respectively, have the chance to capture modal frequencies in the range 0-0.5 Hz, but in fact they have failed. This is because the displacement induced by vehicle loads is always dominated by the static and quasi-static components while the dynamic component of displacement is relatively small (i.e. the root mean square of de-trended acceleration signal during this time interval is only 0.0016 g) and easily contaminated by the measurement noise. It indicates that the displacement sensor (either DMS or GPS) has the capacity to capture the dominant frequency component but might fail to capture some frequency components lower than the Nyquist frequency.

3.3 Field test on Tamar Bridge

Tamar Bridge spans the River Tamar between the City of Plymouth on the east bank and the town of Saltash on the west bank. A single day of field testing using the DMS was performed to measure the lateral and vertical displacement at mid-span of the bridge. The system performance is evaluated by comparing with GPS measurement.

3.3.1 Test configuration

As at Humber Bridge, the hardware included a GigE camera equipped with a 300 mm lens, data acquisition system and a (smaller) artificial target. The camera was installed at the top of a steel tower shown in Fig. 3.5(a), 380 m away from mid-span. The data acquisition system including the controller and a monitor in Fig. 3.5(c) was set at the nearby office of the bridge maintenance team. A custom-made 750 mm square target frame was mounted on the parapet at mid-span indicated in Fig. 3.5(b). Fig. 3.5(d) shows one sample frame captured from the video files recorded on the test day. The derived output was the mid-span displacement in the vertical and lateral directions.

To evaluate the performance of the optical system, an independent GPS system (TOPCON GR-5 RTK) was used in the test. The base station was mounted on a sheltered and stable surveying tripod near the Tamar Bridge office while the GPS rover was attached to the top of the target frame at mid-span in Fig. 3.5(b). Sample rates for vision-based system and GPS were 10 Hz and 2 Hz, respectively.

Firstly, the direct measurement by the vision-based system is evaluated by comparing with the GPS measurement and the possible reasons for differences are discussed. Next, the direct measurement by the vision-based system is corrected according to the tracked ‘motion’ of bridge tower and the corrected result is again compared with the GPS measurement. Note that GPS is not assumed to provide the true displacement signal, but rather a means of comparison.

3.3.2 Measurement results

Ten-minute signals of the DMS and GPS measurement (between 11:20 am and 11:30 am) in the vertical direction are shown with label ‘DMS(Target)’ and ‘GPS’ in Fig. 3.6(a). The signals are artificially offset in the vertical direction by 35 mm to provide a clear view. It is expected that the two measurement signals have the exact same movement shape and amplitude, however, the DMS measurement includes what appear to be high-frequency ‘vibrations’ (e.g. from 11:22 to 11:24) and sharp peaks (e.g. 11.26) that do not appear in the GPS measurement.

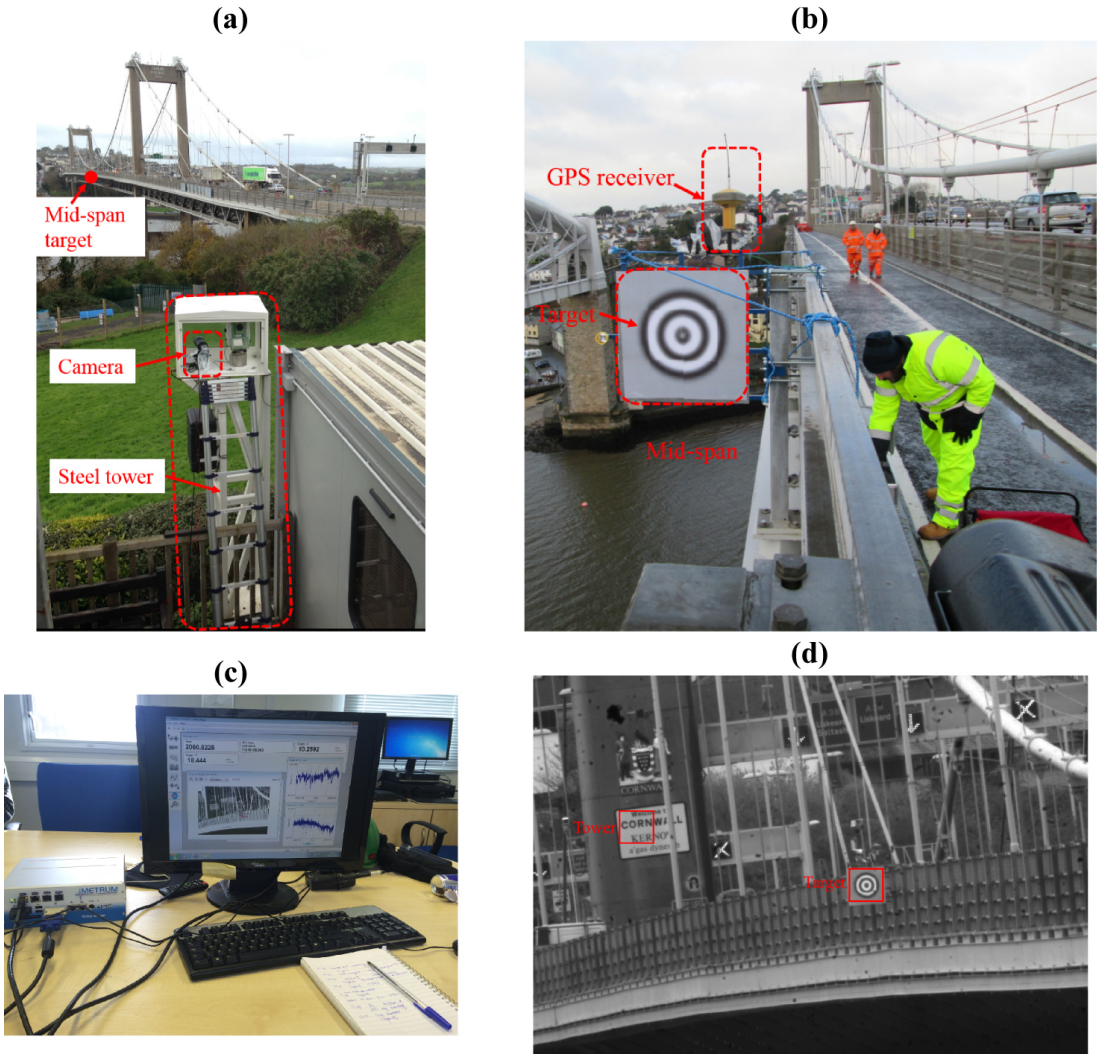


Fig. 3.5 Configuration of Imetrum DMS and GPS: (a) Imetrum DMS; (b) mounting configuration of artificial target frame at mid-span and one GPS rover; (c) data acquisition system for Imetrum DMS; and (d) one sample frame from video files recorded by Imetrum DMS.

The main working principle of vision-based system is by tracking the location of target projection in image and transforming the target locations in image to the true locations in structure via a projection transformation. Thus the error of displacement measurement is mainly induced by the error of the target tracking results and the estimated projection transformation. The target tracking accuracy varies from 0.5 to 0.01 pixel (Bing et al., 2006) which is related to target pattern (texture contrast) (Busca et al., 2014) and illumination condition. The target region in the recorded video file keeps in high-contrast pattern and experiences no sharp lighting change.

The projection transformation from the image plane to the structure system is dependent on camera internal features as well as camera-to-structure position and orientation. Since the camera is fixed in a stable location, the projection transformation is usually determined according to the initial condition without updating as time varies.

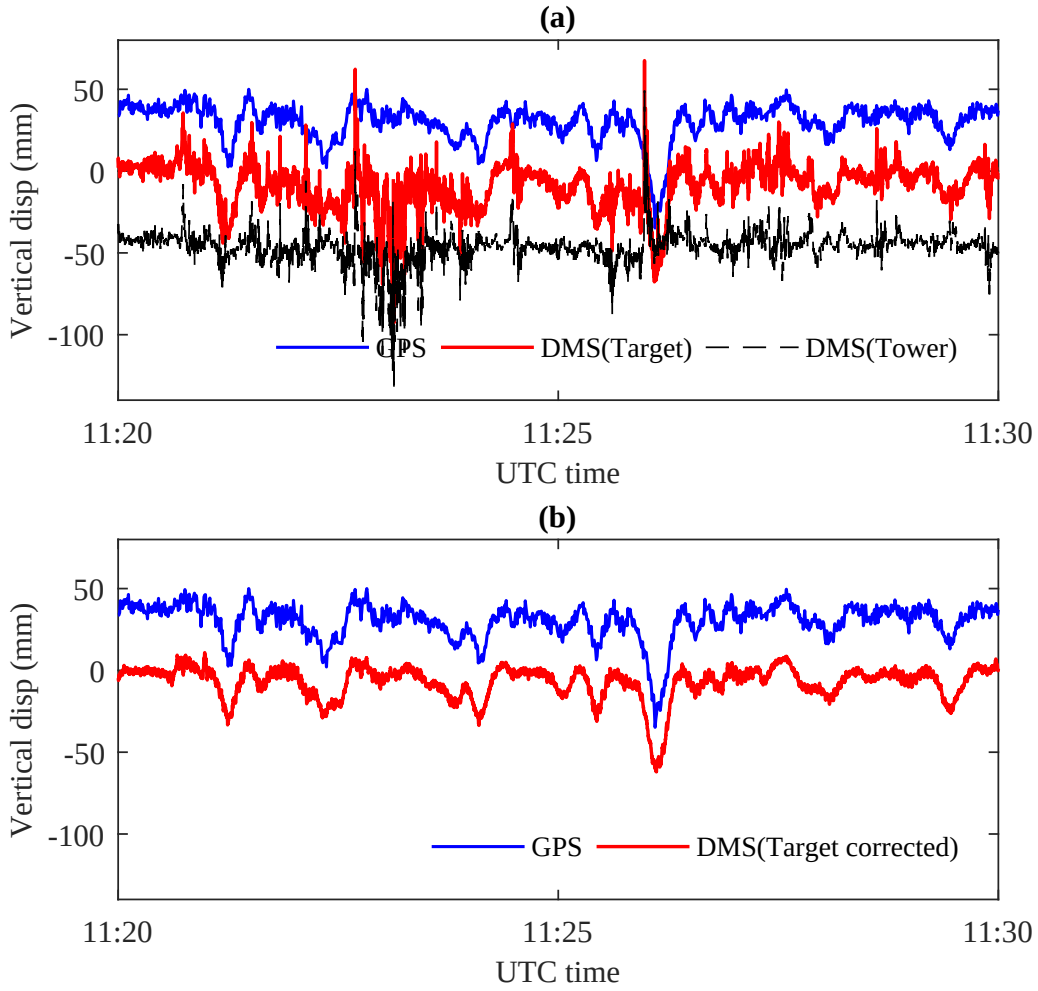


Fig. 3.6 Time-history signals of vertical displacement by GPS and DMS: (a) GPS measurement and DMS raw measurement; (b) GPS measurement and DMS measurement after camera motion correction.

In this test, the camera was mounted at the top of a very stiff steel tower, although a ladder had to be used for setup adjustment. The camera had consistent motion together with the steel tower, possibly leading to error in the DMS measurement when the ladder was used. As a result the raw measurement using the DMS could not be trusted and it was necessary to use correction techniques to compensate for the error induced by the camera motion.

A feasible approach for camera motion correction is through tracking a reference region that is physically fixed; in this case the far (Saltash) bridge tower. The towers were constructed from reinforced concrete and sit on caisson foundations founded on rock (Koo et al., 2013) so are extremely stiff and experience only minute vertical deformations induced by extreme traffic loads (Westgate et al., 2014) and temperature variation (Koo et al., 2013). Hence the true tower deformation would be dominated by low-frequency components small in amplitude compared with the mid-span displacement. The natural feature target in bridge

tower was a rectangular region of a sign ('Welcome to CORNWALL') on the tower surface shown as a red box in Fig. 3.5(d) and tracked simultaneously with the mid-span target. The estimated displacement of the bridge tower shown as a dashed line in Fig. 3.6(a) is not constant but includes the same high-frequency 'vibrations' (e.g. from 11:22 to 11:24) and sharp peaks (e.g. 11:26) observed in the target displacement obtained using the DMS. This proved the assumption that the main DMS error was due to the camera motion. The correction to the directly tracked target location in the image plane is deduced using the tower 'motion' in the image plane and then transformed to structural displacement via a projection transformation (planar homography matrix here).

The corrected result of structural displacement is shown as the lower line in Fig. 3.6(b) together with the GPS measurement. The two signals resemble each other, and the cross correlation coefficient of two signals is calculated to be 95%. Different GPS units were used at Humber Bridge and Tamar Bridge and it seems that the temporary system used at Tamar Bridge provided a more reliable reference than the system used at Humber Bridge.

3.4 Discussion about Imetrum DMS

The Imetrum DMS is validated to be a mature, accurate and stable optical system for bridge deformation measurement over long-ranges and over several hours within a day, but it is a proprietary solution and there remain open various research routes to wider applications and lower costs in non-contact sensing.

In the field applications at Humber Bridge and Tamar Bridge, one single GigE camera with 300 mm lens was used for recording. During the video processing, correlation-based template matching (Potter and Setchell, 2014) was used for target tracking and planar homography method was used for camera calibration to determine the transform relation between the image plane and the structural system. The final output was the 2D structural displacement along the vertical and transverse directions at bridge mid-span.

Regarding the hardware component of a vision-based system:

- Professional high-resolution cameras equipped with long focal length lens used in the DMS are necessary for long-range monitoring. For short-distance monitoring, consumer-grade cameras or smartphones could be alternatives reducing system cost. Consumer-grade cameras have been validated as feasible for displacement measurement and system identification in laboratory testing (Yoon et al., 2016) while reports of field implementations are rare.

- Custom-made artificial targets were used in the study reported here, requiring direct access to the bridge for installation. The role of artificial targets

here includes: i) providing dimensional information for calibrating projection transformation in camera calibration step and ii) providing salient features to improve tracking accuracy. The target tracking algorithm used in the DMS is effective to track the feature target as validated by Fig. 3.3 thus the second function could be ignored. The remaining obstacle of avoiding the need for a cumbersome artificial target is involved in camera calibration step.

Regarding the video-processing methodologies:

- Camera calibration is aimed at determining the projection transformation between the image natural units (pixels) and the real world units (e.g. mm). The scaling factor method using the camera-to-target distance (Khuc and Catbas, 2017a; Yoon et al., 2016) or merging optical system with a total station (Charalampous et al., 2015; Ehrhart and Lienhart, 2015c) has no requirement about known geometric information. However, these applications are based on the prerequisite that the camera principal axis is perpendicular to the target surface plane and are thus not suggested. The general form of projection transformation (i.e. planar homography matrix or full projection matrix) is related to camera internal features (i.e. focal length, principal point locations, etc.) as well as the camera-to-target geometric relation (i.e. position and orientation of camera in structural coordinate system). Parameters describing camera internal features (i.e. camera intrinsic matrix) can be determined in the laboratory ahead of a field test i.e. using the camera to observe a planar calibration object in a few different views (Zhang, 2000). However, determination of camera position and orientation (i.e. camera extrinsic matrix) requires some geometric information from the structure (at least four coplanar point locations). The points with known locations are usually provided by attached artificial targets e.g. planar chessboard target (Chang and Ji, 2007), planar T-shape wand with active markers attached (Park et al., 2015) or 3D calibration object with four non-planar active targets (Martins et al., 2015). To achieve complete non-contact sensing, efforts should be spent on alternative means to obtain dimension information in structure e.g. through surveying or structural design drawings.

- Correlation-based template matching is used for target tracking in the DMS and is validated to provide good performance in either short-range or long-range monitoring campaigns. Correlation-based template matching is based on matching two subset images by similarity level. It assumes that each pixel within the selected rectangular region (target projection) in an image has identical image motion in two-dimensional translation. Thus, template matching is not the best choice to track slender structural components e.g. cables in a cable-supported structure. This is because the target region

bounded by a rectangle window might include some background pixels e.g. clouds and tree branches. The background pixels with motion inconsistent with the true structural motion will contaminate the template matching results. In this case, the preferred choices of target tracking algorithms are optical flow estimation (Yoon et al., 2016) and feature point matching (Khuc and Catbas, 2017a) which are based on tracking sparse points within the target region and are supportive to remove outlier among point correspondences..

In the next section, a custom-developed vision-based system is proposed for multi-point displacement measurement and evaluated in laboratory and field testing. Compared with this commercial product, the proposed system has flexible options of data acquisition devices (e.g. consumer-grade and professional cameras) and offers three advanced tracking approaches to fit different test conditions.

Chapter 4

Performance evaluation of custom-developed vision-based systems

In this chapter, a custom-developed vision-based system is proposed for displacement monitoring on bridge structures. About the hardware, the system has flexible options of video acquisition devices and supports either consumer-grade or professional cameras. In terms of algorithms, the video processing package offers three advanced tracking methods all feasible for natural patterns.

The working performance of the proposed system is evaluated in laboratory and field testing conditions, respectively. Firstly, a system consisting of a consumer-grade camera (GoPro) is tested in laboratory to evaluate the influence of target pattern distinctiveness and tracking methods used on measurement accuracy.

Then, a system consisting of a professional camera is implemented for monitoring tests on one short-span railway bridge and one long-span road bridge. Several field challenges are involved in these two tests including long-range sensing, low-contrast target patterns, pattern changes and lighting changes. The working performance of the three tracking methods in undesired test conditions is compared, indicating their relative advantages and disadvantages and offering some suggestions about tracking method selection in field testing. Discussions about error sources and field limitations of vision-based monitoring are presented in the end.

The field testing on a long-span bridge (i.e. the Humber Bridge) in this chapter is the same test as reported in Chapter 3. The difference is that in this chapter, the commercial product Imetrum system is used only as a video acquisition device. The video records are analysed using the custom-developed video processing package.

This chapter is reproduced from the paper titled "**Vision-based systems for structural deformation measurement: case studies**", co-authored with James

Brownjohn, which was published in *Proceedings of the Institution of Civil Engineers - Structures and Buildings* (Xu and Brownjohn, 2018b).

4.1 Introduction

Vision-based monitoring methods have promising features e.g. simple instrumentation and installation, operation remote from the structure and capacity for multi-point measurement using a single (camera) sensor. Existing studies have indicated the potential of vision-based methods for structural condition assessment, in particular for system identification (Caetano et al., 2007; Oh et al., 2015; Yoon et al., 2016), finite element model calibration (Feng and Feng, 2015), damage detection (Cha et al., 2017) and contribution to bridge WIM system with camera assistance for traffic monitoring (Ojio et al., 2016).

Target tracking is one critical component in a vision-based system, directly influencing the measurement accuracy. Template matching (Brownjohn et al., 2017; Chang and Ji, 2007; Ehrhart and Lienhart, 2015c; Fukuda et al., 2013; Guo and Zhu, 2016; Macdonald et al., 1997; Stephen et al., 1993) and optical flow estimation (Caetano et al., 2011; Cha et al., 2017; Chen et al., 2015a, 2017, 2015b; Diamond et al., 2017; Ehrhart and Lienhart, 2015a; Ji and Chang, 2008a; Khaloo and Lattanzi, 2017; Yang et al., 2017; Yoon et al., 2016) are established methods widely used for bridge deformation measurement whereas feature point matching is a relatively new and promising tracking method that is theoretically scale-invariant and rotation-invariant (Ehrhart and Lienhart, 2015c) and has been validated in several short-range measurement tests (Ehrhart and Lienhart, 2015a,c; Khuc and Catbas, 2017a,b). There are also some other methods through tracking the special shapes of target patterns based on edge detection or image thresholding algorithms (Ji and Chang, 2008b; Lee et al., 2006; Ribeiro et al., 2014; Wahbeh et al., 2003) e.g. line-like cables, circular-shaped dots and chessboard, etc. These methods have limitations for application due to the requirement about pattern shapes.

To find the most appropriate tracking method for structural monitoring, it is necessary to evaluate their measurement accuracy and the robustness to different field conditions. Busca et al. (2014) evaluated three techniques (template matching, edge detection and digital image correlation) on a steel truss railway bridge, concluding that the three techniques provide similar tracking performance while tracking accuracy is slightly poorer for natural targets. Ehrhart and Lienhart (2015c) evaluated the performance of three techniques (optical flow, template matching and feature point matching) by tracking structural features of a footbridge and reported that feature point matching is robust to the changes of background condition (i.e.

snowfall) whereas drift over time was observed in the measurement by two other methods.

These two existing studies provide some information about the influence of pattern features and pattern changes on measurement results. However, more studies are necessary to evaluate the field performance of vision-based systems since several critical field challenges are not considered yet e.g. robustness to lighting changes and viability for long-range monitoring.

The purpose of this study is to investigate the effectiveness of vision-based systems for displacement measurement in different environmental conditions through three case studies. One laboratory test and two field tests were performed indicating several influential factors on measurement performance, i.e. estimation error in projection transformation, camera-to-target distance, the distinctiveness of target patterns, changes of target patterns and changes in lighting conditions. Three representative tracking methods were considered for video processing, demonstrating their advantages and shortcomings to deal with the observed influential factors. Other error sources in field tests like camera shake, atmospheric refraction and temperature variations were not apparent in test observations and thus are not discussed in this study.

This work is the first study to apply the feature point matching method in the long-range monitoring test and the measurement was evaluated through comparison with the processing results using classical tracking methods. The field measurement by vision-based systems might carry some low-frequency error due to camera shake or lighting changes. A method to distinguish the main source of low frequency error is proposed in this study.

To that end, section 4.2 introduces the methodologies for vision-based displacement measurement, in particular, three representative tracking methods. In section 4.3, a laboratory uniaxial oscillation test used to validate the video processing methods and to study the tracking accuracy in laboratory conditions is described. Section 4.4 and 4.5 report the field tests for mid-span deformation measurement in a short-span railway bridge and a long-span suspension bridge, respectively. The performance of vision-based systems with challenging field conditions e.g. low-contrast patterns, changes of target patterns and changes in lighting conditions are investigated, indicating the advantages and shortcomings of the three tracking methods. Finally the main findings from the three tests are summarised.

4.2 Vision-based systems

In vision-based systems, the hardware comprises data acquisition devices, portable computer with video processing algorithms and accessories like tripod and artificial

targets (optional). Video processing package is the key part which could fit into a three-component framework in Fig. 4.1, i.e. camera calibration, target tracking and displacement calculation.

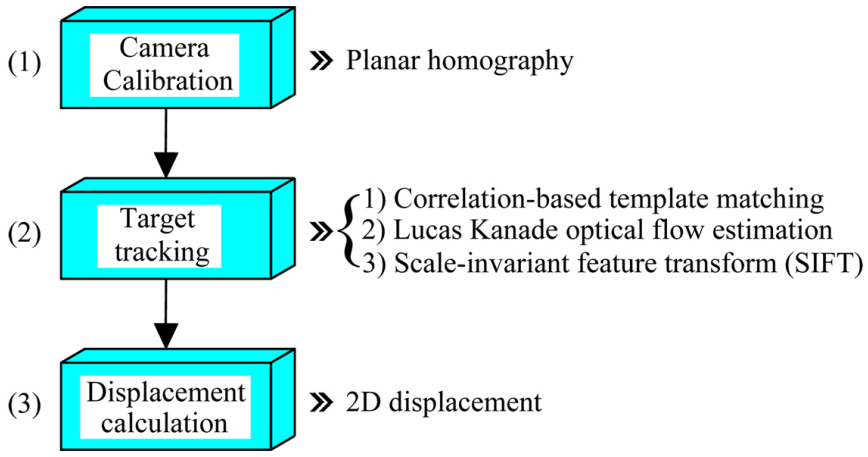


Fig. 4.1 Procedures and methodologies in custom-developed video processing package for structural displacement measurement.

The custom-developed vision-based system in this study consists of a camera (either consumer-grade or professional) and a post-processing package for extracting displacement information from video records. The package is developed using C++ language, partly referring to OpenCV library. This custom-developed package is capable to measure the structural displacement of multiple targets simultaneously within the field of view and offers three options of tracking methods (shown in Fig. 4.1) to adapt to different test conditions.

The video processing package consists of four classes defined for the realisation of camera intrinsic calibration, extrinsic calibration, target tracking and structural displacement calculation, respectively. The package is easy to use: parameter settings are pre-specified in a text file as the input (see Fig. 4.2, an example for the class of extrinsic calibration) and output data are written to a text file with a pre-defined format after running.

Target tracking is critical in the video processing package to locate the target regions in the image plane through tracking methods. The tracking methods considered in this study include template matching and optical flow estimation that are established and classic, as well as feature point matching that receives increased attention in structural monitoring.

This section provides a description of the methodologies of three target tracking methods used in the custom-developed video-processing package i.e. correlation-based template matching, Lucas Kanade (LK) optical flow estimation and scale-invariant feature transform (SIFT) method (Lowe, 2004).

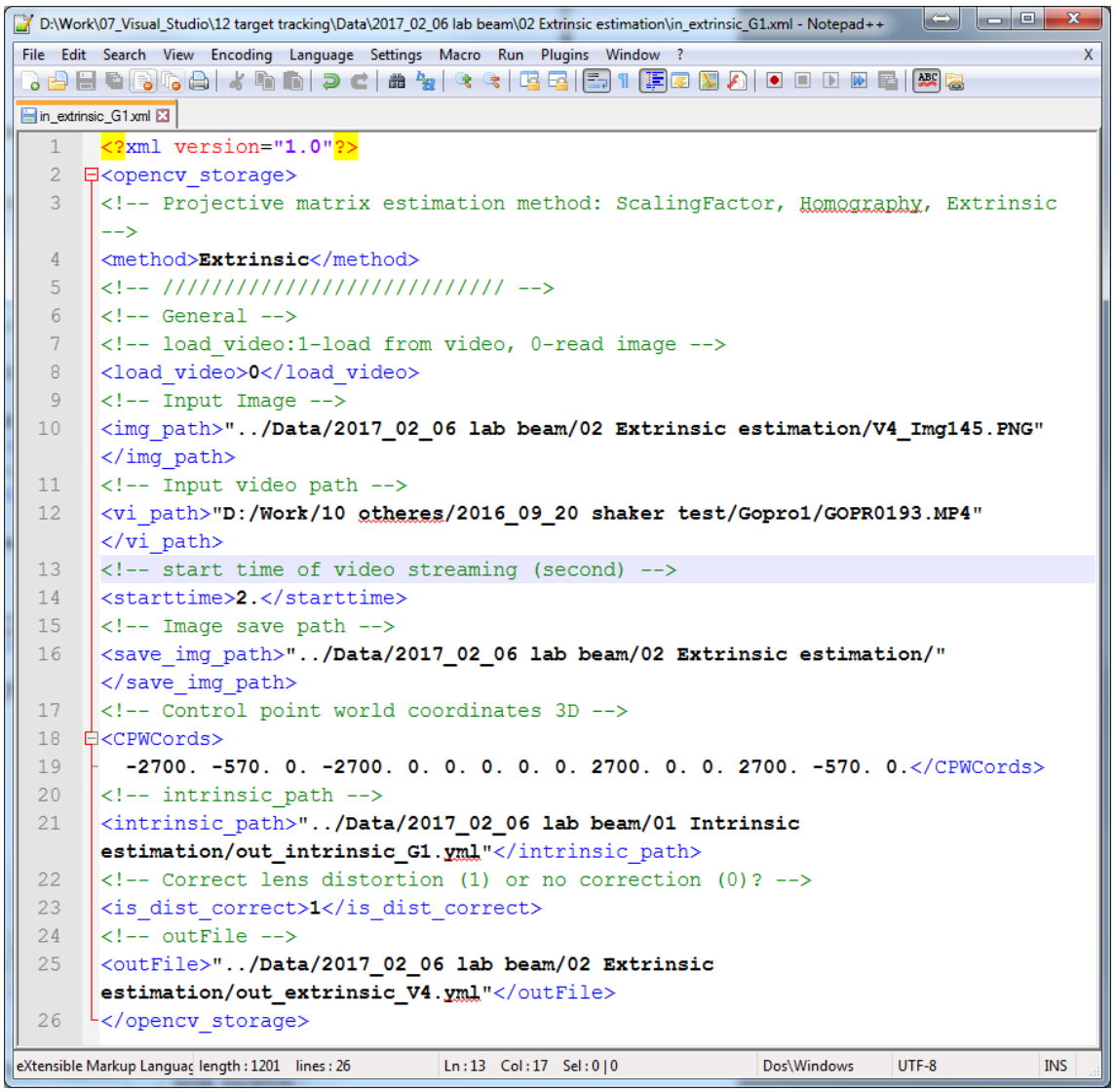


Fig. 4.2 A input file sample in the video processing package.

4.2.1 Correlation-based template matching

Template matching is a classic technique for target tracking by searching in a new frame for an area most closely resembling a predefined template, following the procedures demonstrated in Fig. 4.3. A target region is selected as the template that is a subset image in the reference frame. A matching criterion is defined to evaluate the similarity degree between the template and the new frame and the criterion used is zero-mean normalised cross correlation coefficient (ZNCC). The target location in the new frame corresponds to the peak location in the similarity matrix that has resolution at pixel level. Subpixel interpolation schemes (Feng et al., 2015a) are required to refine the tracking results to sub-pixel level and the interpolation method used in this study is zero-padding in frequency domain using the matrix multiplication form of discrete Fourier transform (Guizar-Sicairos et al., 2008).

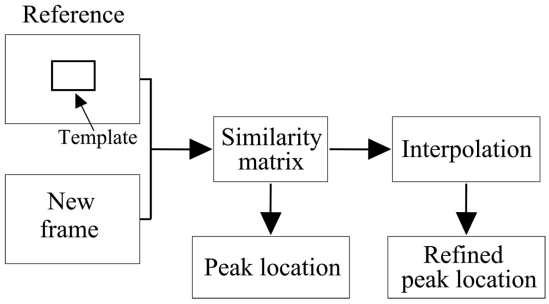


Fig. 4.3 Procedures of one target tracking method: correlation-based template matching.

Template matching has been applied in structural monitoring since the earliest work on the Humber and Severn Bridges in 1990s (Macdonald et al., 1997; Stephen et al., 1993), and the recent applications include displacement monitoring tests in a railway bridge (Feng et al., 2015a), a long-span bridge (Ye et al., 2013) and a high-rise building (Liao et al., 2010).

4.2.2 Lucas Kanade optical flow estimation

Lucas Kanade optical flow estimation detects the motions in an image from the brightness pattern shift (Beauchemin and Barron, 1995). The calculation process imposes one temporal constraint on image properties and one spatial constraint on motion consistency, i.e. that the pixel intensities of an object do not change between consecutive frames and that neighbouring pixels have similar motion. The image motions are derived using the following equation,

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} \sum_i I_{xi}^2 & \sum_i I_{xi} I_{yi} \\ \sum_i I_{xi} I_{yi} & \sum_i I_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_{xi} I_{ti} \\ -\sum_i I_{yi} I_{ti} \end{bmatrix} \tag{4.1}$$

where dx and dy denote the motions in image plane; I_x, I_y and I_t represent the spatial and temporal gradients of image intensities; and i denotes the ith pixel location in a 3×3 patch around a feature point (x, y).

The procedures are demonstrated in Fig. 4.4. The most prominent corners are detected by Shi-Tomasi corner detector (Shi et al., 1994) in the reference frame with their locations updated in the new frame by Equation 4.1. Backward estimation is then performed from the new frame to the reference frame in order to refine the point correspondences based on the error between the initially detected points and the backward-estimated points. Point correspondences are further refined based on geometric alignment using the least median of squares method (Massart et al., 1986).

The LK optical flow estimation was applied in a laboratory test of a multi-storey metal tower (Yoon et al., 2016) for system identification, for field application of a

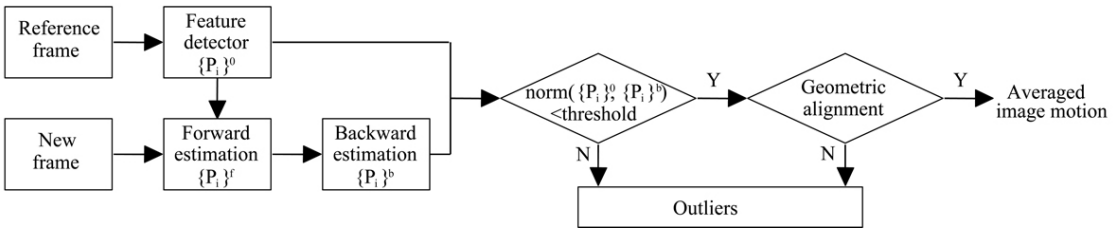


Fig. 4.4 Procedures of one target tracking method: Lucas Kanade (LK) optical flow estimation.

footbridge deformation monitoring (Ehrhart and Lienhart, 2015a) and for bridge stay-cable vibration measurement (Caetano et al., 2011; Ji and Chang, 2008a).

4.2.3 Scale-invariant feature transform (SIFT)

Feature point matching is an efficient tracking technique through matching feature points in consecutive images based on their local appearance. Several robust feature extractors and descriptors are reported in literature (Alahi et al., 2012; Bay et al., 2008; Calonder et al., 2010; Lowe, 2004; Rublee et al., 2011) and the one used in this study is the SIFT method.

The procedures are demonstrated in Fig. 4.5. Keypoints are extracted from the local extremes in the Difference of Gaussian images (differences of Gaussian filtered images with varied blur level) and described by vectors using the gradient magnitudes and orientations of neighbouring pixels. Keypoints between two images are matched by identifying their nearest neighbours evaluated using the Euclidean distances between keypoint descriptor vectors. The outliers in keypoint correspondences are removed based on geometric alignment using the least median of squares method (Massart et al., 1986).

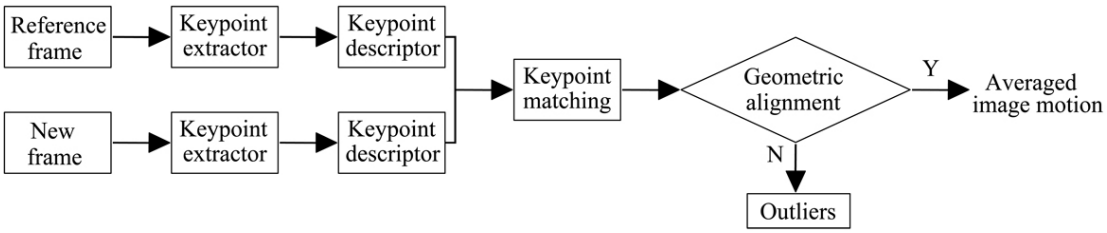


Fig. 4.5 Procedures of one target tracking method: scale-invariant feature transform (SIFT).

The SIFT method has been validated in the deformation measurement test of a railway bridge (Khuc and Catbas, 2017b). Other feature point matching methods are also validated only in short-range monitoring tests (Ehrhart and Lienhart, 2015a,c; Khuc and Catbas, 2017a).

4.3 Validation test in controlled environment

This section investigates the performance of three tracking methods (i.e. correlation-based template matching, LK optical flow estimation and SIFT method) in controlled environmental conditions through a laboratory uniaxial oscillation test.

4.3.1 Test configuration

An APS 400 Electrodynamic shaker was set vertically on solid ground shown in Fig. 4.6(a) and driven by an input of chirp signal with frequency range 0.5 Hz to 2 Hz. The test was run twice, with no artificial target in Run 1 and with the chessboard pattern attached to shaker mass (shown in Fig. 4.6(b)) in Run 2 in order to increase the feature salience.

Two GoPro Hero4 cameras were fixed on the ground looking at the oscillating shaker mass; sample frames in two runs are given in (c) and (d). The frame rates were set as nominally 60 Hz with narrow field-of-view option. A GoPro Smart Remote was used to control the two cameras simultaneously. A Balluff micropulse transducer was attached to the shaker mass shown in (b) to provide an accurate reference displacement measurement sampled at 256 Hz.

The recorded video files were post-processed using the custom-developed video processing package to derive the displacement data of shaker mass. Camera calibration was performed according to the dimensions of shaker mass indicated in Fig. 4.6(b) to estimate the planar homography or full projection matrix. The target locations in the image were estimated by the three tracking methods, respectively. The displacement data were acquired using the results from the previous two steps. For single camera application, the 2D structural displacement is derived uniquely through transforming the target location in image plane to structural displacement via the planar homography. When two cameras with overlapped views are used to monitor the same target, the 3D structural displacement is extracted using triangulation method (Hartley and Zisserman, 2003).

Section 4.3.2 presents the measurement results by single camera (GoPro camera 1) while Section 4.3.3 provides the results using stereo vision from two camera records.

4.3.2 Measurement results using single camera

The measurement results using the GoPro camera 1 and Balluff transducer in two runs are illustrated in Fig. 4.7, indicating that,

- Vision-based system based on a consumer-grade camera is qualified to measure the vertical oscillation of shaker mass by tracking either natural

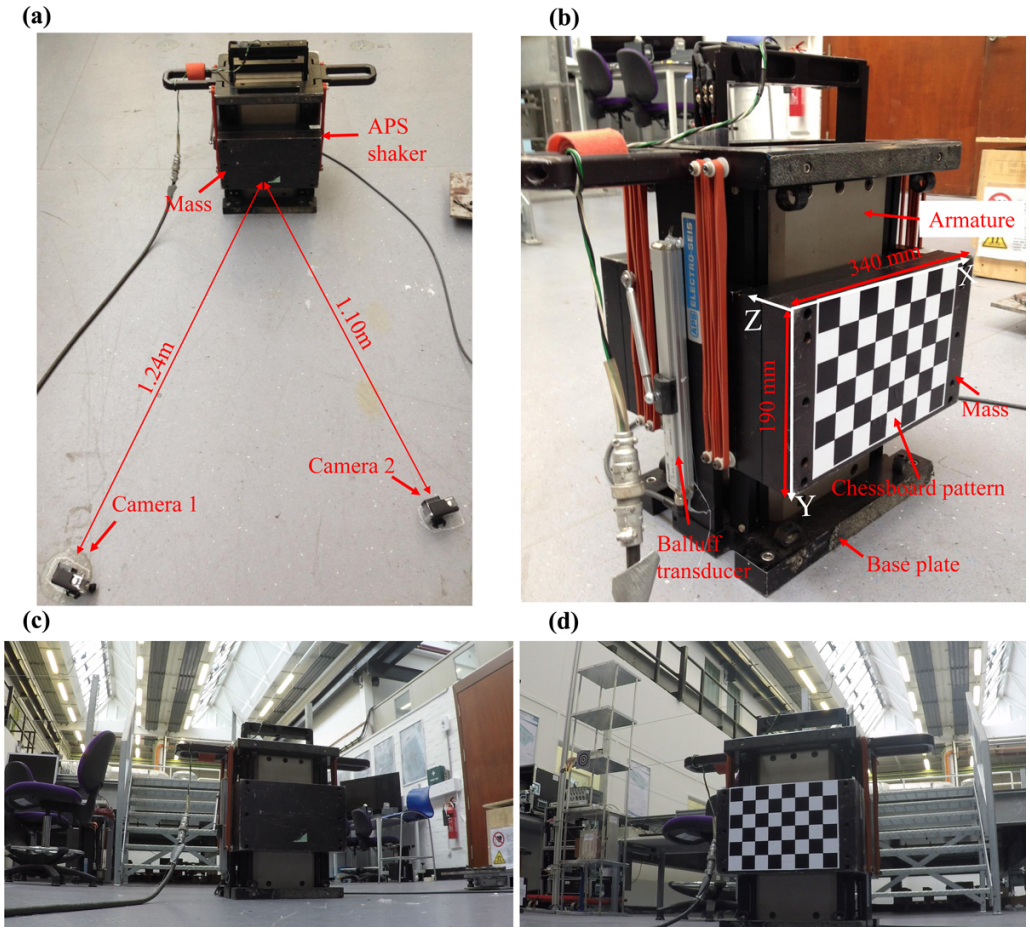


Fig. 4.6 Test configuration of a vision-based system for vertical oscillation measurement of an APS shaker in laboratory: (a) test configuration of shaker and two GoPro cameras in Run 1; (b) test configuration of shaker in Run 2 with chessboard patterns attached to shaker mass; (c) one sample frame from the video recorded by GoPro camera 1 in Run 1; and (d) one sample frame from the video recorded by GoPro camera 2 in Run 2.

features or artificial target patterns with the cross-correlation coefficient (compared with the Balluff measurement) reaching over 99%.

- Oscillation occurred only in the vertical direction and any displacement measurement in the horizontal direction corresponds to measurement error. In Run 1, correlation-based template matching provides the best measurement performance while the SIFT measurement includes the highest noise. Measurement error in Run 2 is much smaller than in Run 1 after increasing the target salience.

- During the oscillation period in Run 2, the horizontal measurement indicates a chirp signal with a similar shape as the vertical measurement but with much smaller amplitude. It might be caused by the error in projection transformation, making the dominant vibration leak to minor motion direction.

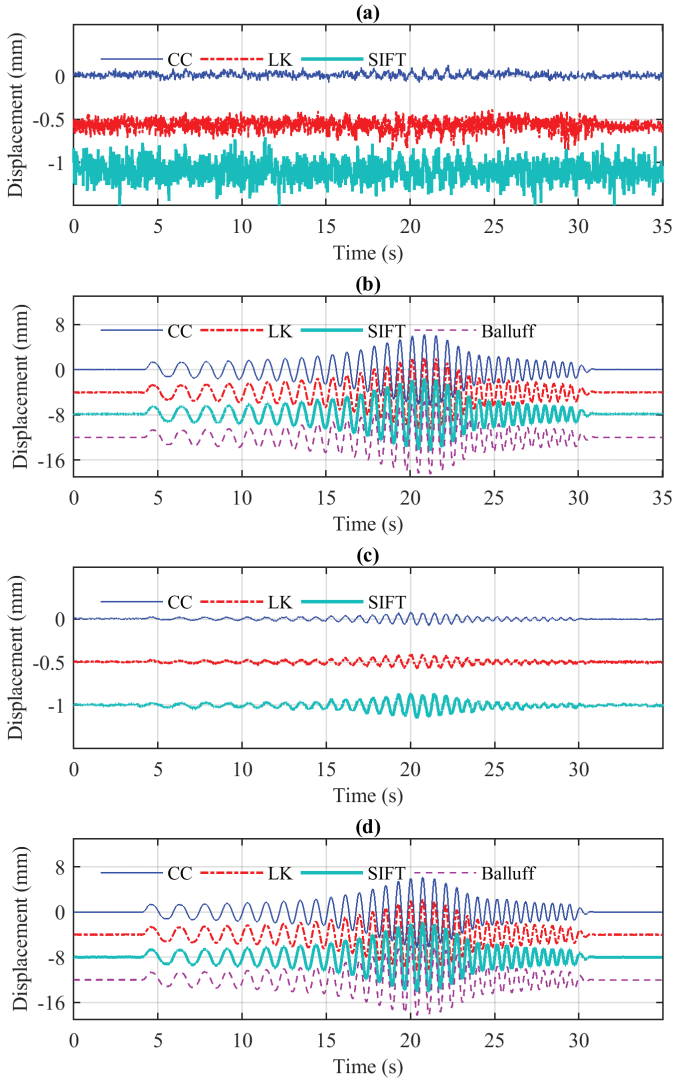


Fig. 4.7 Time histories of displacement measurement of shaker mass acquired by the custom-developed video processing package: (a) measured displacement in the horizontal direction by vision-based system in Run 1; (b) measured displacement in the vertical direction by vision-based system and Balluff transducer in Run 1; (c) measured displacement in the horizontal direction by vision-based system in Run 2; and (d) measured displacement in the vertical direction by vision-based system and Balluff transducer in Run 2. (Legends CC, LK and SIFT denote the three target tracking methods used for video processing, namely correlation-based template matching, Lucas Kanda optical flow estimation and Scale-invariant feature transform.)

To investigate the tracking accuracy of the three methods, the data collected during the stationary period were taken into account with the estimation results summarised in Table 4.1. This indicates that in laboratory conditions, correlation-based template matching is the most accurate method while the SIFT achieves

the poorest accuracy with the highest sensitivity to the distinctiveness of target patterns.

Table 4.1 Tracking accuracy at 95% confidence interval for three tracking methods in laboratory conditions

Tracking accuracy (95% confidence interval)	Correlation-based template matching	LK optical flow estimation	SIFT
Run 1 (pixel)	0.026±0.055	0.029±0.069	0.048±0.149
Run 2 (pixel)	-0.009±0.020	-0.001±0.020	0.011±0.037
Accuracy range (pixel)	[0.01, 0.08]	[0.02, 0.10]	[0.03, 0.20]

4.3.3 Measurement results using stereo vision

Two Gopro cameras were implemented to reconstruct the shaker mass oscillation together. Although a GoPro remote control was used for time synchronisation in recording process, an apparent signal lag between two measurements (about 0.05 s) is observed in the post-processing process indicated in Fig. 4.8. The signal lag was estimated by finding the maximum cross correlation coefficient of two displacement signals using single camera calibration.

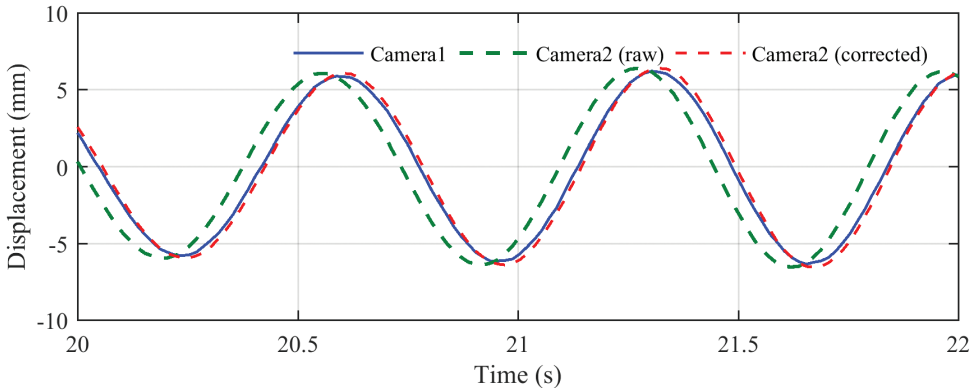


Fig. 4.8 Time synchronisation for two camera records

The 3D displacement is reconstructed using the calibration information from the two cameras. Fig. 4.9 presents the measurement error in Run 1 (using correlation-based template matching method) evaluated by the reference sensor. Results indicate that the measurement error is very slightly reduced in horizontal direction after applying stereo vision. For single camera sensing, the largest error in the vertical direction appears at the periods (21∼23 s) when the oscillation amplitude is largest. This error in vertical direction is decreased using stereo reconstruction.

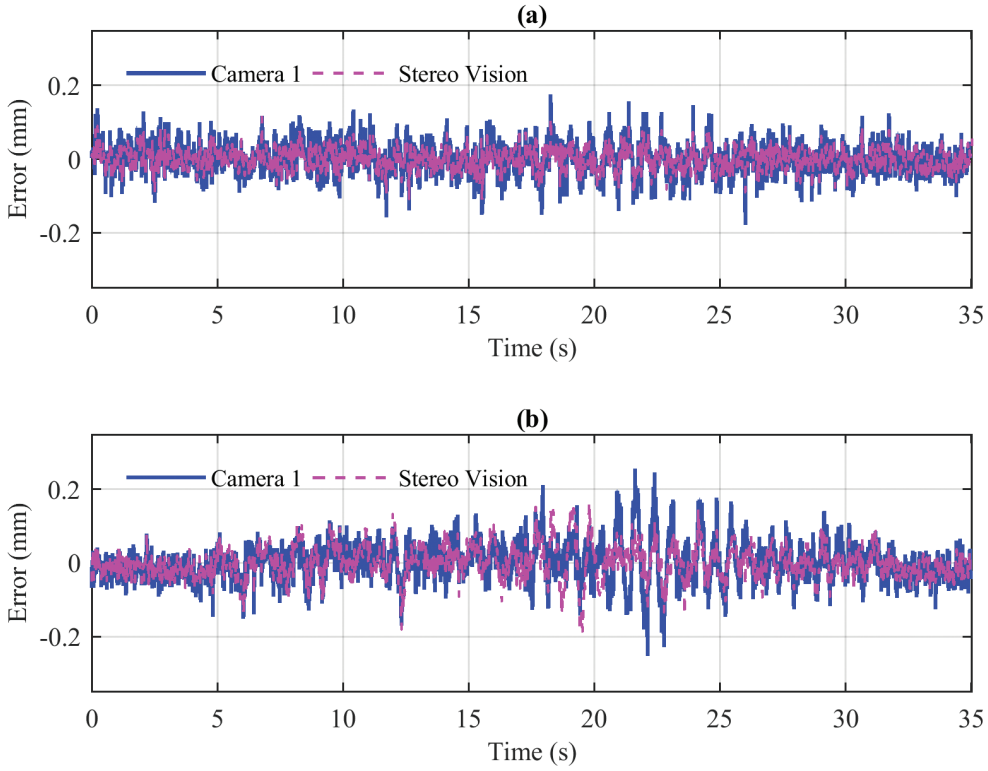


Fig. 4.9 Measurement error of displacement data using stereo reconstruction in horizontal (a) and vertical directions (b)

The comparison shows that stereo reconstruction could bring in some (not overwhelming) improvement over single-view monitoring provided two or multiple cameras are perfectly synchronised.

Although a stereo configuration contributes more reliable results than a single-view arrangement, it requires higher installation efforts and an accurate time-synchronisation of camera recordings. Instead of using multiple-view arrangement, the single-view configuration applying planar homography method for camera calibration is implemented on field tests discussed in this thesis. This configuration has no constraint on camera positioning but assumes that the out-of-plane motion is negligible. The assumption could be feasible for most bridge applications since deformation in the vertical direction is dominant compared with those in two horizontal (longitudinal and transverse) directions. The measured output in field tests is time-histories of two-dimensional displacement of bridges along vertical and transverse directions or along vertical and longitudinal directions.

Laboratory evaluation avoids numerous difficulties that can diminish performance in field conditions due to environmental influences such as camera-to-target distance, unstable target patterns and lighting conditions, etc. In section 4.4 and 4.5, two field tests on bridge deformation measurement are reported to illustrate the real-world working performance of vision-based systems.

4.4 Deformation measurement test on a short-span bridge

The vision-based system applied in the two field tests is the Imetrum Dynamic Monitoring Station (DMS) originating from research at the University of Bristol and commercialised via the university spin-out company Imetrum formed in 2003. The system includes one or more GigE high performance cameras for data acquisition and the software 'Video Gauge' for the real-time video processing.

Target tracking algorithms used in the software are proprietary extensions of correlation-based template matching techniques which enable better than $1/100$ pixel resolution at sample rates beyond 100 Hz in field applications. The system used by University of Exeter has been trialled in several one-day field campaigns on a number of bridges in the UK, indicating comparable or even better measurement accuracy compared to an LVDT for short-range measurement,(Hester et al., 2017) and the GPS for long-range measurement (Xu et al., 2017).

In this study, the Imetrum system was used in field tests mainly for two functions, 1) as a data acquisition device to record the video files of the bridges that would be analysed using the custom-developed video processing package, and 2) to provide (using the Imetrum proprietary video processing software) the reference data of bridge deformation for the evaluation of the measurement results provided by the custom-developed video processing package.

This section reports a case study using the vision-based system for the deformation measurement of a short-span railway bridge. Section 4.4.1 and 4.4.2 below describe the test setup and the results obtained, respectively.

4.4.1 Test configuration

The Mineral Line Bridge is a steel girder bridge with 14.68 m span, carrying the West Somerset Railway near Watchet, UK. Fig. 4.10(a) indicates the test setup with an Imetrum camera mounted at the top of a tripod approximately 12.5 m away from the mid-span of the bridge. One sample frame captured by the Imetrum camera is illustrated in Fig. 4.10(b) with the target region for measurement marked by a rectangular box.

Since the fundamental frequency of the bridge was estimated to be approximately 10 Hz, a frame rate of 30 Hz was set for the Imetrum system. The camera calibration was performed using the dimensions of the bridge girder and the artificial target at mid-span. The displacement data along the vertical and longitudinal directions was measured directly by the Imetrum system and also extracted from the custom-developed post-processing package using the three tracking methods.

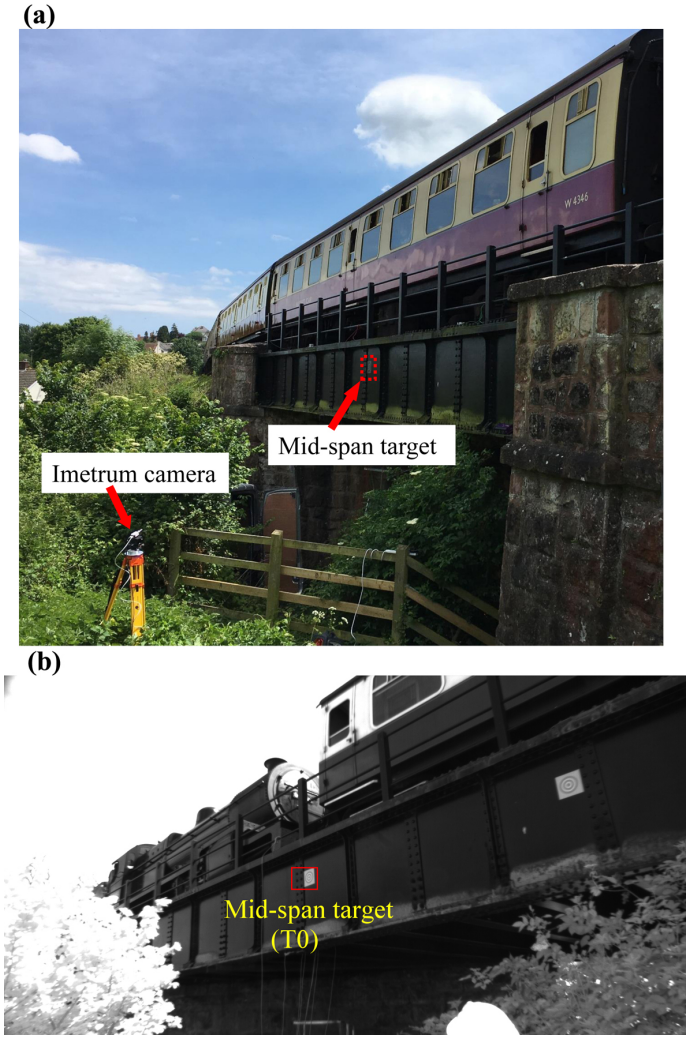


Fig. 4.10 Test configuration of a vision-based system for the mid-span displacement measurement on a railway bridge in Somerset, UK: (a) camera setup near the bridge and the target region at mid-span selected for video tracking; and (b) one sample frame from the recorded video when one steam train passed through the bridge.)

4.4.2 Measurement results

Time history data of displacement measurement are illustrated in Fig. 4.11, indicating that,

- The three tracking methods all capture the vertical deformation of the bridge induced by the passing of a steam train with one steam locomotive and seven passenger carriages with the maximum deformation at 6.4 mm.
- A similar deformation pattern during the train passing occurs in the longitudinal measurement with much smaller amplitudes (less than 0.56 mm) apart from the measurement by SIFT method due to the high noise. It might be caused by the error of projection transformation, making the vertical deflection data leaked to the horizontal direction.
- An apparent low-frequency motion trend in the horizontal direction is observed after 30 s for all the four measurement that should be an error since the bridge was empty with no heavy loading. During this period, the four methods provide different amplitude values and the SIFT measurement is the most noisy one but with the smallest amplitude. The LK optical flow estimation method failed to measure in some frames possibly due to the large brightness changes. In terms of the image motion, the maximum drift in the image horizontal direction reaches approx. 0.3 pixel, larger than the estimated tracking accuracy in laboratory conditions. It indicates that the tracking accuracy becomes poorer in field conditions for any of the three tracking methods.

According to the authors' test experiences and the literature, the low-frequency error in vision-based measurement could be caused by either the camera motion (Zhao et al., 2017) or changes in lighting conditions (Brownjohn et al., 2017). The error induced by camera motion should be consistent at those pixels corresponding to the stationary objects in the field of view. The camera motion is believed not to be the main error source in this case because Fig. 4.11 is for one among several targets tracked in the bridge girder (in Fig. 4.12(a)), and these show image motions inconsistent in both the amplitude and direction as shown in Fig. 4.12(b) and (c).

To quantify the influence of lighting changes, mean pixel intensity at the initially selected target region (T0) were calculated as shown in Fig. 4.13(a). This indicates a growth of averaged brightness from 28.5 s to 37 s followed by a gentle decrease, which has a trend similar to the measurement error in the longitudinal direction in Fig. 4.11(a). The initial frame and the frame at the time step 37 s are shown in Fig. 4.13(b) and (c) for visual comparison of lighting changes. In the second frame, the bolts within the rectangular region (T0) are more distinctive against the background due to the improved lighting. Therefore, it is believed that the low-frequency error in the longitudinal direction is caused by the lighting changes.

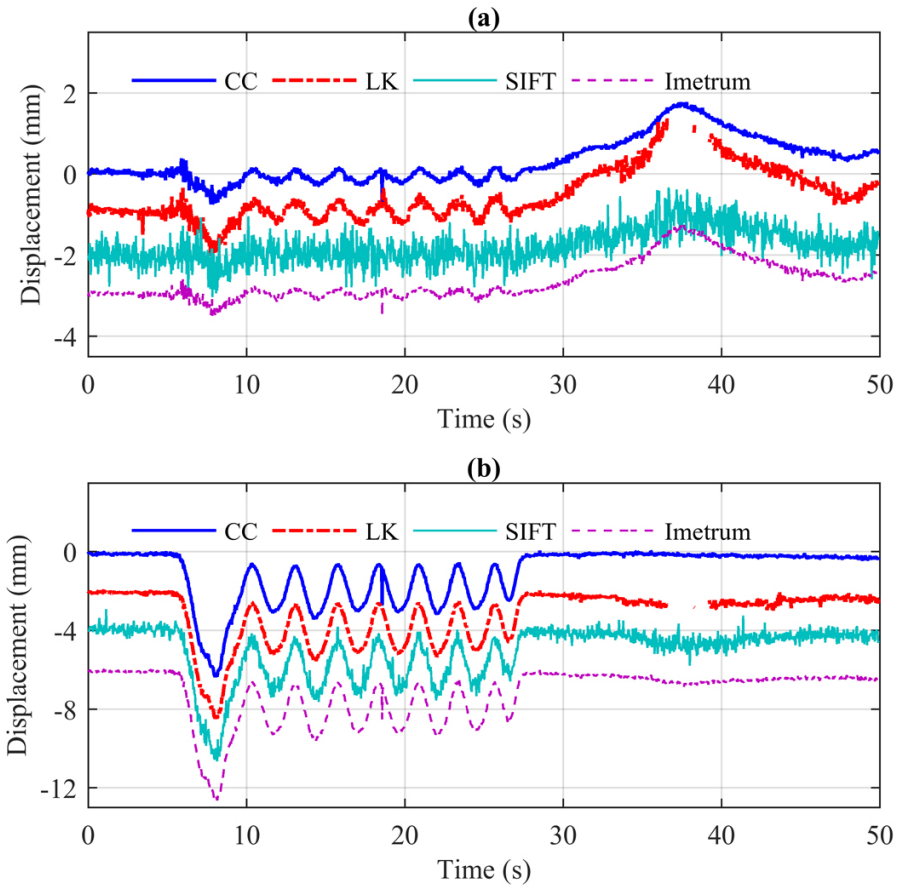


Fig. 4.11 Time histories of displacement measurement acquired by the Imetrum system and by the custom-developed video processing package using three tracking methods: (a) displacement measurement in the bridge longitudinal direction; and (b) displacement measurement in the vertical direction. (Legends CC, LK and SIFT denote the three target tracking methods used for video processing, namely correlation-based template matching, Lucas Kanda optical flow estimation and Scale-invariant feature transform.)

None of the three tracking methods is robust to large lighting changes in field tests. The SIFT method experiences the least influence in measurement while the LK optical flow might fail to identify the features under apparent lighting changes. The low-frequency error due to lighting changes might mislead the users about the bridge loading condition if no prior knowledge is available.

A sharp brightness increase is observed at time step 18.57 s in Fig. 4.13(a) corresponding to an outlier in the measurement by correlation-based template matching and the Imetrum system in Fig. 4.11. The cause of these anomalies is changes in target pattern due to a bird flying in front of the target, as shown in Fig. 4.14. This indicates that the LK optical flow estimation and SIFT method are not sensitive to small pattern changes whereas the correlation-based template matching method might fail to track.

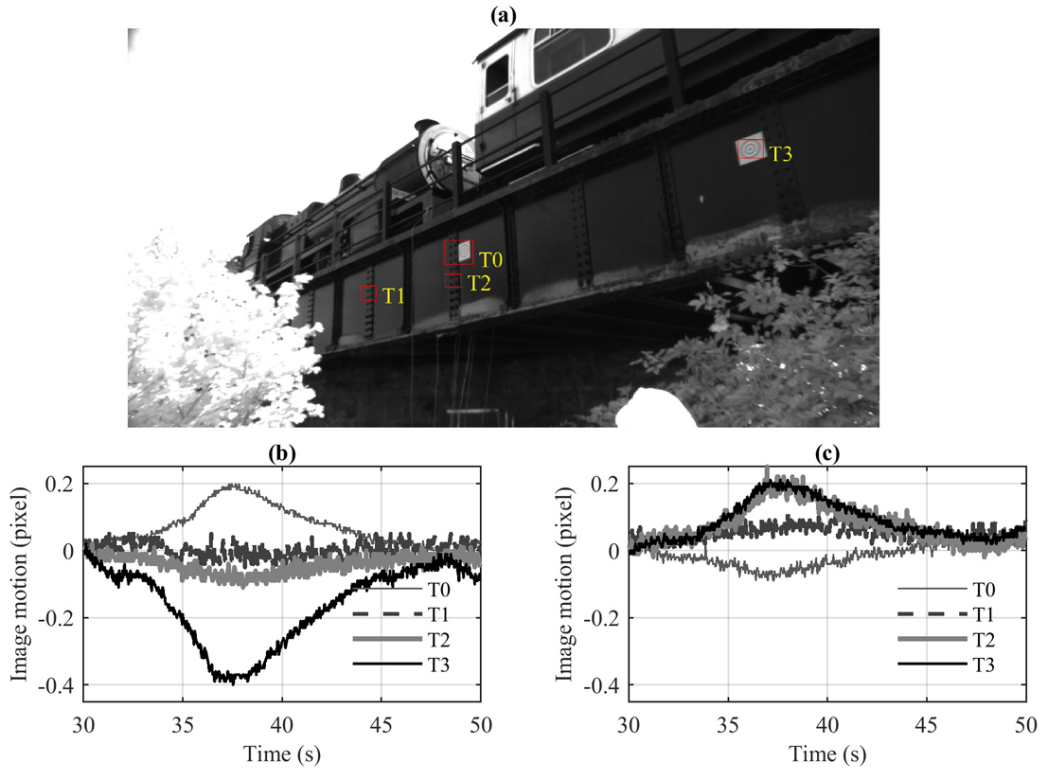


Fig. 4.12 Tracking results of image motions at the selected four target regions (T0-T3) in the bridge girder by the correlation-based template matching: (a) locations of four target regions in the frame selected for tracking; (b) measured image motions at four target regions along the image horizontal direction; and (c) measured image motions at four target regions along the image vertical direction.

The demonstration of vision-based monitoring in a short-span bridge indicates that the vision-based system is capable to measure the bridge deflection under traffic loads using any of the three tracking methods, although the measurement might become unstable when suffering from apparent changes in lighting conditions or target patterns. In the next section, a similar test was performed on a long-span bridge to investigate the viability of vision-based system for the long-range monitoring.

4.5 Deformation measurement test on a long-span bridge

This section describes a case study of using a vision-based system measuring deformation of the Humber Bridge. The test has been reported in Chapter 3.2.1 to evaluate the working performance of a commercial vision-based system Imetrum. Here, the Imetrum system is only taken as a video acquisition device and the video files are post-processed using the custom-developed package. The purpose

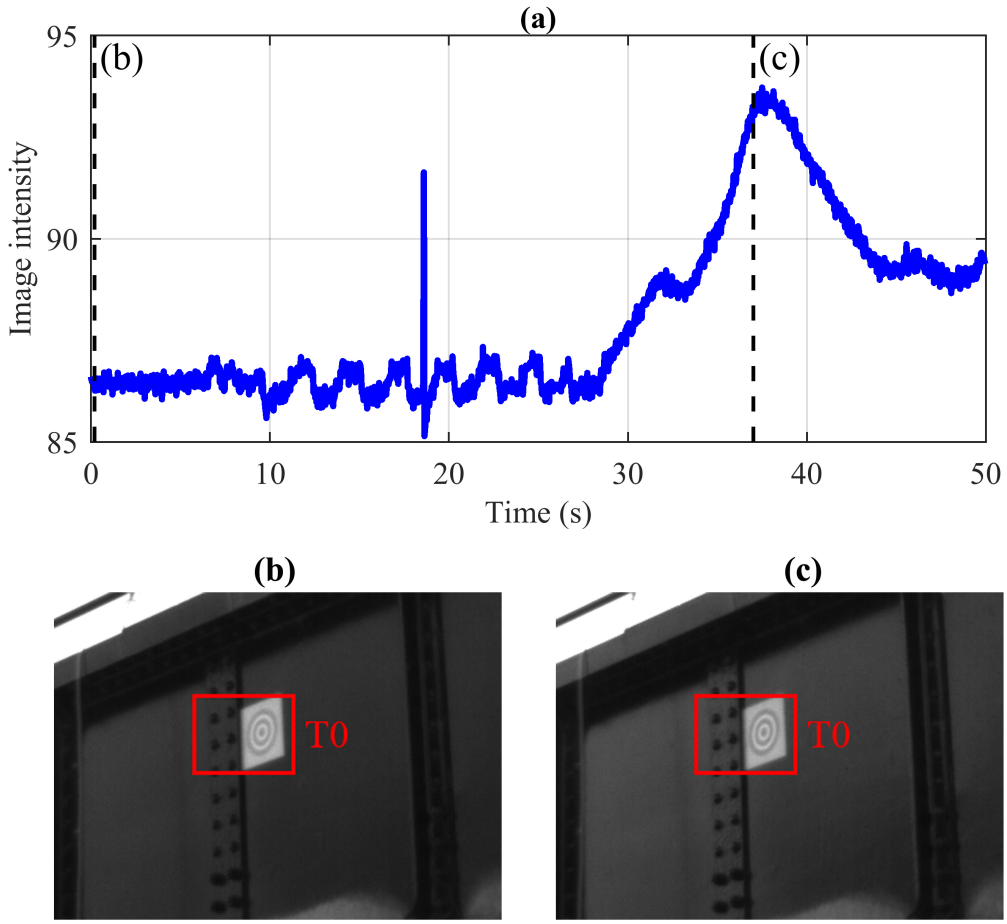


Fig. 4.13 Variations of image brightness at the initially selected target region T0: (a) time history of mean pixel intensity at the target region T0; (b) the truncated initial frame with a rectangular annotation at the target region T0; and (c) the truncated frame at the time step of 37 s with a rectangular annotation at the target region T0.

of re-analysing video records is for the performance evaluation of three tracking methods for the long-range measurement.

Measurement results are presented directly in this section while the test configuration could refer to Fig. 3.2.

An 80-second signal of vertical displacement at the target region T1 recorded at approx. 19.22 PM (BST) is illustrated in Fig. 4.15. The maximum displacement reaches 160.5 mm at approximately 31 s by the Imetrum measurement while some data points using the three post-processing methods are missing for about 0.9 s when the displacement values reach their maxima. Two frames during this period are shown in Fig. 4.16 indicating a big change in target pattern. Due to the low sun elevation in the west, the target panel on the east side was initially partially in the shadow of the bridge railing, shown in (a). When one tall vehicle passed the mid-span of the bridge between the sun and the target, sunlight was completely blocked, making the whole target pattern visible in the image. This indicates that the three tracking methods are all not robust to large changes in target patterns. Imetrum system, with its proprietary algorithms, is more robust in this case.

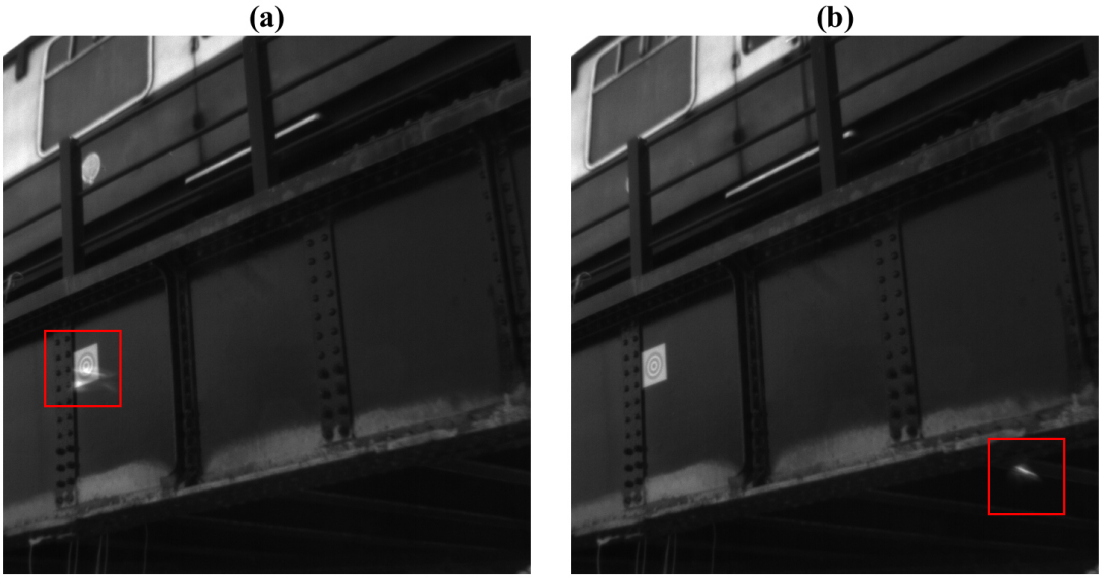


Fig. 4.14 Two consecutive frames from video files at approx. 18.6 s indicating the changes of target pattern due to a flying object (frames truncated and zoomed-in for clarification).

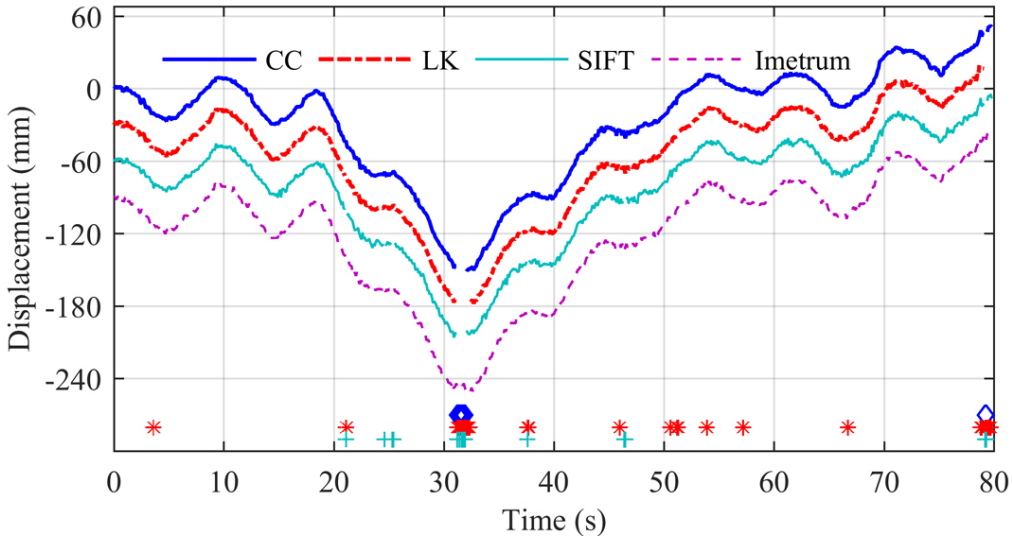


Fig. 4.15 Time histories of displacement measurement in the vertical direction at the target region T1 acquired by Imetrum system and by the custom-developed video processing package using three tracking methods with the markers indicating the time steps of tracking failure. (Legends have the same meaning as in Fig. 4.11.)

The target region T2 is located at the deck soffit that is less salient with smaller spatial changes in target patterns compared with the target region T1. In Fig. 4.17, the maximum displacement measured at T2 is 128.2 mm and 130.7 mm by the correlation-based template matching and SIFT method, respectively while the LK optical flow method failed to track several frames including the period reaching the maximum displacement. This indicates that the LK optical flow method has higher requirements on salience and stability of target patterns. The displacement

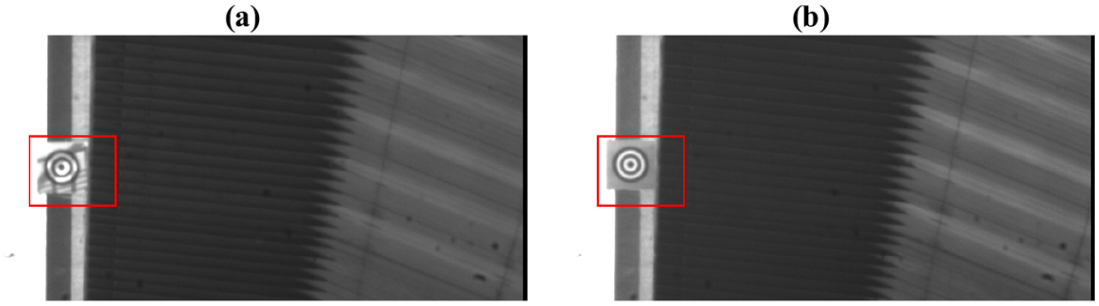


Fig. 4.16 Two frames from the video file at approx. 31.3 s indicating the changes of target pattern due to the passing of one tall vehicle at the mid-span of the bridge that temporally blocked the sunlight, making the whole target pattern visible in the image.

measured at T2 is smaller than that measured at T1 because 1) the two targets would experience different motion as the bridge rotates about its longitudinal axis due to eccentric traffic loading; and 2) the bridge axis directions projected in the image plane and the projection transformation were determined according to the artificial target panel and thus were not perfectly aligned with those for the feature target T2 under the deck soffit.

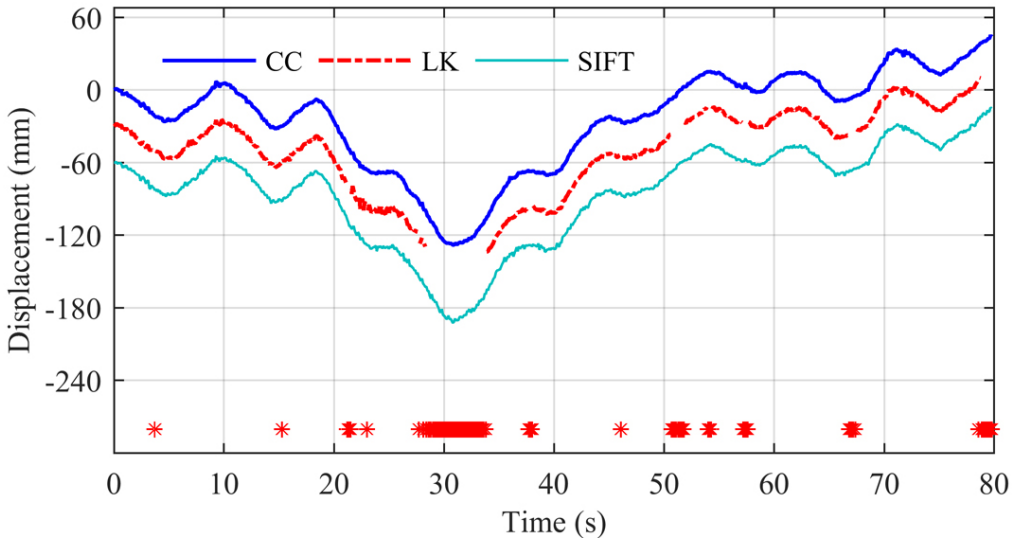


Fig. 4.17 Time histories of displacement measurement in the vertical direction at the target region T2 acquired by the custom-developed video processing package using three tracking methods with the markers indicating the time steps of tracking failure.

4.6 Conclusion

This chapter investigates the performance of vision-based systems for displacement measurement in laboratory and field tests. Three representative tracking

methods (i.e. correlation-based template matching, LK optical flow estimation and SIFT method) were used for the post processing of test records with their performance compared with a commercial vision-based system, Imetrum DMS.

In laboratory conditions, the tracking accuracies for the two methods, correlation-based template matching and LK optical flow estimation are close varied from 0.02 pixel to 0.10 pixel depending on target patterns while the accuracy of SIFT method is poorer in the range between 0.03 pixel and 0.20 pixel.

The working performance of three tracking methods in field tests are summarised in Table 4.2.

Table 4.2 Working performance of three tracking methods in field tests

Field conditions	Correlation-based template matching	LK optical flow estimation	SIFT
Long-range measurement	✓	✓	✓
Low-contrast patterns	✓	×	✓
Small pattern changes	×	✓	✓
Large pattern changes	×	×	×
Lighting changes	×	×	×

- All the three tracking methods are effective for either short range or long range measurement (e.g. camera-to-target distance at 710 m) with the displacement varying from several millimetres to ten centimetres. However, the tracking accuracy becomes poorer than that achieved in laboratory conditions.

- The salience of target patterns has a direct influence on the measurement accuracy and high-contrast patterns are preferred for tracking. LK optical flow estimation has the highest requirement about the distinctiveness and stability of target patterns and might fail to track when the other two methods work fine.

- Changes to target patterns due to object obstruction or daytime shadows might lead to missing data. Correlation-based template matching is the method most sensitive to the changes of target patterns while the other two tracking methods are also influenced when facing large changes on target patterns.

- Changes of lighting conditions might cause some low-frequency measurement error using any of the three tracking methods, which could be misunderstood without the prior knowledge of structural loading. SIFT method is influenced by the lighting changes but provides smaller measurement error compared with the other two methods.

It is indicated that apart from restraining the test conditions, e.g. testing in overcast weather or selecting the sheltered target patterns, how to deal with varying lighting conditions in the field is still an open question for vision-based measurement.

Another important observation is that although the two-dimensional displacement measurement is provided by the vision-based system, the measurement along the minor deformed direction might not be reliable. This is because the error in projection transformation might lead to the leakage of dominant deformation to the minor deformed direction. Thus special attention should be given to interpret the measured displacement along the minor deformed direction.

Chapter 5

Enhanced data fusion for displacement and acceleration measurement

As demonstrated in Chapter 4, vision-based monitoring faces several field challenges and might fail to achieve desired accuracy level in some real-world test conditions e.g. involving strongly varying ambient lighting, low-contrast patterns and instability of optical instrument, etc. To make the best use of the potential of vision-based systems, an effective way is data fusion which allows the integration of data from a series of sensors to make a better estimation than that from any sensor alone.

Efforts have been made to integrate displacement data with other sensors (commonly accelerometers) for accuracy measurement. However, most of the existing methods in literature involve supervised learning process or users' judgement for the selection of certain 'parameters'.

This chapter proposes a practical and autonomous data fusion method to merge displacement data with the collocated acceleration for accuracy improvement. The main algorithm is based on multi-rate Kalman filter assisted with the maximum likelihood estimation (MLE) that is implemented to determine the necessary noise parameters automatically. The MLE enhanced data fusion method is validated on numerical and field data, indicating the capacity of improving displacement accuracy and widening frequency bandwidth.

The field testing data is from a single-day displacement monitoring campaign on the Humber Bridge, same as reported in Chapter 3. The demonstration example here is to improve the direct measurement by GPS through fusing with the collocated acceleration data. Vision-based system is taken as a reference sensor in this chapter since the environmental conditions on the test day were optimal for camera monitoring (calm day and stable lighting conditions).

This chapter is focusing on development and validation of the proposed data fusion method while implementing the data fusion process on vision-based systems is presented in in Chapter 6.

This chapter is reproduced from the paper titled "**Long-Span Bridges: Enhanced data fusion of GPS displacement and deck accelerations**", co-authored with James Brownjohn, David Hester and Ki Koo, which was published in *Engineering Structures* (Xu et al., 2017).

5.1 Introduction

Currently the global positioning system (GPS) is basis for the common sensing of displacement in long-span bridges (Brownjohn et al., 2015). However, the accuracy level of GPS data is a major concern. Nickitopoulou et al. (2006) suggest standard accuracy of up to 15 mm and 35 mm for horizontal and vertical measurements, respectively, at 98.5 percentile level without gross errors such as cycle slip or multipath. Casciati and Fuggini (2009) confirmed the accuracy to be of the order of sub-centimetre level through both static and dynamic tests. Recently high-rate GPS receivers e.g. up to 100 Hz are available due to the developments in GNSS receiver and antenna technologies (Yi et al., 2013b), but the increase of sample rate may not provide any improvement, and may even degrade accuracy (Moschas and Stiros, 2014). These tests were conducted in an open area with no reflective surface whereas for the field implementations especially in long-span cable-stayed or suspension bridges, the measurement accuracy becomes worse since the multipath effects near cables and deck are more obvious (Nickitopoulou et al., 2006).

5.1.1 Previous efforts to improve GPS accuracy

An effective way to improve the GPS measurement accuracy is data fusion which allows the integration of data from a series of sensors to make a better estimation than that from any sensor alone (Sohn et al., 2002). Efforts have been made to integrate the GPS measurement with other sensors (commonly accelerometers). Li et al. (2006) demonstrated the complementary feature of GPS and accelerometer measurement in low-frequency and high-frequency ranges through analysing the in-site measurement of a tall building during the earthquake and typhoon. Roberts et al. (2004) proposed a sensor configuration to ensure the time synchronisation and measurement axis alignment between two types of sensors. The post processing methods for data fusion include: 1) integrating acceleration signals to displacement while correcting the result using GPS-derived initial condition (Roberts et al., 2004), 2) reconstructing displacement using the low-frequency

component derived from GPS measurement and high-frequency component derived from acceleration measurement (Chan et al., 2006), and 3) denoising GPS measurement (filter) based on the dominant frequencies derived from acceleration signals (Moschas and Stiros, 2011; Yi et al., 2013c). Another effort is linking the GPS measurement with Robotic Total Station (RTS) measurement through superimposition of semi-static displacement from RTS and dynamic displacement from GPS (Moschas et al., 2013) or a supervised learning approach to distinguish the oscillation amplitude from the measurement noise (Psimoulis and Stiros, 2012). These approaches are supervised learning process with filter or correction and thus require the users' judgement and involvement for the selection of certain 'parameters', e.g. determining the cut-off frequency for signal denoising or selecting the effective frequency bandwidth of measured acceleration and displacement signals for time-history superimposition.

A promising approach to link autonomously GPS measurement with acceleration signals is to use Kalman filter, an approach that has been actively applied in the fields of navigation, computer vision and economics, etc. In the structural monitoring field, Smyth and Wu (2007) proposed a multi-rate Kalman filter approach to fuse the measured acceleration and displacement signals at different sample rates while Chang and Xiao (2010a) applied the method to a pedestrian bridge by integrating the data from the accelerometer and a vision-based system. Kim et al. (2014) modified this fusion method by adding acceleration measurement bias into state variables. In the Kalman filter based fusion methods, there is no displacement base correction nor any frequency-selective filter for noise reduction. Instead, the Kalman filter method finds the best estimate at each time step by weighted average between the noisy displacement measurement and the recursive prediction made from previous best estimate. This estimation process requires reasonably accurate models of the noise in both displacement measurement and state transition model (e.g. measurement and process noise covariances). However, these noise covariances are assumed as known parameters in the abovementioned works, which in many situations is not the case. Therefore, Li and Chang (2013) proposed an adaptive subspace technique to quantify time-sensitive variances of noise embedded in measurement signals whereas this requires knowledge of the structure dynamic characteristics and multi-point measurement data, which again is not always available.

5.1.2 Focus of this chapter

From the above text it is evident that challenges remain on how to measure the displacement in long-span bridges using GPS with suitable measurement accuracy and acquisition frequency. It has been shown previously that a Kalman filter can be used to merge displacement data with corresponding acceleration data to

improve displacement measurement accuracy (Smyth and Wu, 2007)) provided that noise parameters such as process and measurement noise covariances are known. Therefore this study proposes the novel step of using maximum likelihood estimation (MLE) to accurately estimate the noise parameters required by the Kalman filter process.

The effectiveness of the proposed approach is first demonstrated on numerically simulated data of a single degree of freedom system. Afterwards, this method is applied to improve the accuracy of GPS measurement recorded on the Humber Bridge, UK. To examine the effectiveness of the approach when applied to field data, the results are compared to the displacement recorded using an independent vision-based system. Although the vision-based system has some limitations e.g. requirement of good environment and clear view, the system performance on a calm day with stable observation conditions was validated to be satisfactory over several hours in day time through mid-span displacement monitoring tests in long-span bridges including the Vincent Thomas Bridge, USA (Wahbeh et al., 2003), Tsing Ma Bridge, China (Ye et al., 2013) and P25A bridge, Lisbon (Martins et al., 2014).

In the past other researchers have looked at the possibility of ameliorating GPS displacement data by fusing it with corresponding acceleration data, so the main purpose of this chapter is to investigate the improvement in the accuracy of displacement readings that can be achieved by incorporating MLE in data fusion process.

To this end, section 5.2 gives an overview of the data fusion method based on the Kalman filter algorithm and the principles of maximum likelihood estimation while section 5.4 describes the field testing at Humber Bridge, UK and the results observed.

5.2 Data fusion of acceleration and displacement

In effect this chapter proposes to combine the work of Smyth and Wu (2007) with maximum likelihood estimation (MLE), and a brief overview of each is given below.

Smyth and Wu (2007) utilise the multi-rate Kalman filter to fuse the acceleration and displacement measurement with different sample rate for an improved estimation of displacement signals. (Specific details on their approach are provided in section 5.2.1). In this method, users need to specify the noise parameters e.g. process and measurement noise covariances (Q and R respectively in this study). These parameters are related to the measurement noise in measured displacement and acceleration signals. The difficulty is that these noise parameters are in fact unknown and the erroneous selection of noise parameters leads to poor performance of displacement estimates.

MLE is a common method in statistics to estimate the parameters of a statistical model (e.g. mean and standard deviation) from an incomplete sample of observed data i.e. data on the whole population is not available. In this study, MLE is used to estimate the parameters related to process and measurement noise covariances required by the method of Smyth and Wu (2007). Further information on MLE is presented in section 5.2.2 below.

5.2.1 Multi-rate Kalman filter with backward smoothing

Smyth and Wu (2007) developed a method to merge measured displacement data with collocated acceleration data to obtain improved estimation of displacement. The method uses the multi-rate Kalman filter with forward estimation and backward smoothing. Summary details are reproduced here to highlight issues concerning some noise parameters that are required but unknown, and to define a number of variables discussed subsequently.

State variables for the discrete-time state space model are introduced as displacement and velocity:

$$\mathbf{X}(k) = [x(k), \dot{x}(k)]^T \tag{5.1}$$

The dynamic motion can be predicted recursively in a linear state-space representation as,

$$\mathbf{X}(k+1) = \mathsf{A}\mathbf{X}(k) + \mathsf{B}u(k) + w(k) \tag{5.2}$$

and the displacement measurement is modelled as a Gaussian process,

$$z(k) = \mathsf{H}\mathbf{X}(k) + v(k) \tag{5.3}$$

where k denotes the time step; and $u(k)$ and $z(k)$ are the measured data of acceleration and displacement, respectively. The matrices A, B and H are given by

$$\mathsf{A} = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix},\ \mathsf{B} = \begin{bmatrix} dt^2/2 \\ dt \end{bmatrix},\ \mathsf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{5.4}$$

where dt is the sampling time interval of acceleration signals.

The parameter $w(k)$ is the uncertainty in state-transition model or the process noise assumed as a two-dimensional Gaussian process with zero mean and fixed covariance Q. The parameter $v(k)$ denotes the measurement noise assumed as a white noise Gaussian process with zero mean and fixed variance R. The covariance matrices Q and R can be obtained by

$$\mathsf{Q} = \begin{bmatrix} qdt^3/3 & qdt^2/2 \\ qdt^2/2 & qdt \end{bmatrix},\ \mathsf{R} = \frac{r}{dt} \tag{5.5}$$

where q and r denote the variance of acceleration and displacement noise in the analogous continuous-time system. It is noted that the GPS observations are dominated by both the white and the coloured noise (due to systematic errors) (Moschas and Stiros, 2013), but here only the white noise is modelled since the systematic errors during a short duration possibly have nearly constant amplitudes (Psimoulis and Stiros, 2012).

The discrete-time Kalman filter algorithm and backward smoothing technique can be applied for the state estimation (the mean $\hat{\mathbf{X}}$ and the covariance matrix **P** of optimal estimate of state variables) including three main steps, time update, measurement update and backward smoothing. The corresponding details could refer to Smyth and Wu (2007).

During the implementation of Kalman filter, parameters modelling the uncertainty of state-transition model and measurement model (e.g. Q and R or q and r) need to be determined ahead before any state estimation. Obviously, the accuracy level of the state estimation depends on the selection of these parameters, but unfortunately the true value of these variables is typically unknown in practice.

In our state model, the parameters q and r are the noise variances of acceleration and displacement in the continuous-time model. Ideally the noise variances could be estimated by taking some stationary measurement and analysing the acquired data, but the stationary measurement is not always available, especially in a field test. Also the noise level might not be constant with time and space, e.g. the GPS noise level becomes larger in the case of narrower distribution of satellites at certain times of the day.

It is essential to find an appropriate way to determine the noise parameters needed for Kalman filter. In previous work on this topic either the procedure for determining the noise parameters is not given (Kim et al., 2014; Smyth and Wu, 2007) or they are estimated by the energy of noise component separated from the noisy signals using an adaptive subspace technique (Li and Chang, 2013). This subspace method requires the input of multi-channel measurement in which the number of channels is at least equal to the number of identifiable modes, which is neither particularly satisfactory nor practical. Therefore in the next section, the MLE method is presented and is used to tune the noise parameters and initial state of state variable before application of the Kalman filter.

5.2.2 Maximum likelihood estimation (MLE) of noise parameters

The MLE is a common method to estimate the unknown parameters of a statistical model (e.g. mean and standard deviation) when some observational data are available. In essence, the approach works by

- assuming a particular statistical distribution model (e.g. normal distribution) for an event with some unknown statistical parameters,
- collecting observational data about the event, and
- carrying out a parametric study to identify the model parameters (e.g. mean and standard deviation) so that the possibility of getting observed data under this parameter choice is greater than that under any other choice.

Thus the MLE is an optimisation process to find the value of unknown parameters in a statistic model that maximises the likelihood of this statistic model given the observational data. The following equations and text summarise the main mathematical steps in the process. The statistic event here is corresponding to the displacement information at each time step. The statistical model and inherent unknown parameters are provided in Equation 5.6 and 5.7. Equations 5.8 and 5.10 give the calculation formulas for log-likelihood function that will be maximised to derive the optimal estimate of unknown model parameters.

In our application, the observational data $\mathbf{Z}_N$ are a sequence of displacement measurements $z_k(k = 1, 2, ..., N)$. The displacement measurement is a Gaussian process with the probability density derived from the measurement equation 5.3 in the Kalman filter referred to Yuen et al. (2007),

$$\begin{aligned} p(Z_k|\theta) =& (2\pi)^{-1} det\left(\mathsf{H}\mathbf{P}(k|k-1)\mathsf{H}^T + R\right)^{-1/2} \\ & exp\{-\frac{1}{2}[z(k) - \mathsf{H}\hat{\mathbf{X}}(k|k-1)]^T(\mathsf{H}\mathbf{P}(k|k-1)\mathsf{H}^T + R)^{-1} \\ & [z(k) - \mathsf{H}\hat{\mathbf{X}}(k|k-1)]\} \end{aligned} \tag{5.6}$$

The symbol θ represents the unknown parameters in the statistical model of displacement measurement including i) the noise variances q and r of measured acceleration and displacement in the continuous-time system, and ii) the initial state of state variables $\mathbf{X}(0)$. Thus,

$$\theta = [q, r, \mathbf{X}(0)] \tag{5.7}$$

The likelihood of the parameters in a statistical model given the observational data is equal to the probability of those observed data given the model parameters shown in the following equation,

$$l(\theta|\mathbf{Z}_N) = p(\mathbf{Z}_N|\theta) = \sum_{k=1}^{N} p(z_k|\theta) \tag{5.8}$$

Since the natural logarithm of likelihood function or log-likelihood is monotonically related to the likelihood function, the log-likelihood function is deduced using Equations 5.6 and 5.8 for computational convenience,

$$\begin{aligned}\ln(\theta|\mathbf{Z}_N) = &- N\ln(2\pi) - \frac{1}{2}\sum_{k=1}^{N}\ln\, det(\mathbf{H}\mathbf{P}(k|k-1)\mathbf{H}^T + R) \\ &- \frac{1}{2}\sum_{k=1}^{N}[z(k) - \mathbf{H}\hat{\mathbf{X}}(k|k-1)]^T(\mathbf{H}\mathbf{P}(k|k-1)\mathbf{H}^T + R)[z(k) - \mathbf{H}\hat{\mathbf{X}}(k|k-1)]\end{aligned} \tag{5.9}$$

The optimal estimate of unknown parameters $\theta^\star$ is derived by maximising the log-likelihood function using nonlinear optimisation algorithm,

$$\theta^\star = \text{argmax}\ \ln l(\theta|\mathbf{Z}_N) \tag{5.10}$$

The log-likelihood function can have many local maxima. In order to get a global maximum, a large sets of initial model parameters $\Theta^0 = \theta_1^0, \theta_1^0, ..., \theta_M^0$ are generated based on the uniform distribution within a specified lower and upper bounds. For each set of initial vector θ_j^0, an optimum $\theta_j^\star$ is found using the MLE method. By comparing the log-likelihood of all derived optima $\Theta^\star = \theta_1^\star, \theta_1^\star, ..., \theta_M^\star$, the one corresponding to the maximum log-likelihood is taken as the global optimum $\theta^\star$.

The output of the MLE is the optimal estimate of unknown model parameters $\theta^\star$ and will be used directly in the Kalman filter to fuse the displacement and acceleration signals. Multi-rate Kalman filter supplemented by the MLE is expected to provide more accurate estimates of displacement than the isolated multi-rate Kalman filter. In order to evaluate quantitatively the performance of this method, a simulation example of the single degree of freedom (SDOF) oscillation is conducted before the field application.

5.3 Numerical validation on simulated data

The purpose of this section is to access the effectiveness of the proposed data fusion method. This is achieved by examining its performance when applied to simulation data from a single degree of freedom (SDOF) oscillator subjected to an earthquake excitation. The process of applying both MLE and the Kalman filter estimation is demonstrated. Subsequently the results from sole Kalman filter estimation is compared to the results from Kalman filter estimation supplemented by the MLE. Section 5.3.1 generates the displacement and acceleration signals from a numerical model of a SDOF system. These signals are made more representative of those likely to be encountered in the field by adding white noise to the numerically generated acceleration and displacement signals. Section 5.3.2 describes the process of applying the MLE to predict the optimal noise parameters for use in the state space model. Finally section 5.3.3applies the multi-rate Kalman

filter for displacement estimation using several sets of noise parameters and validates that the optimal noise parameters by the MLE provide more accurate displacement estimate than when MLE is not used.

5.3.1 Simulation example of a SDOF system

This section describes a simulated example of a SDOF system under the earthquake excitation. The measured signals of displacement and acceleration were simulated by contaminating the true responses by a white noise process.

A SDOF oscillator is subjected to the recorded ground motion from the El Centro Earthquake. Equation 5.11 gives the dynamic motion formula of the system,

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = -m\ddot{x}_g(t) \tag{5.11}$$

where m is the mass, c is the damping coefficient, and k is the spring stiffness. $\ddot{x}(t)$, $\dot{x}(t)$, $x(t)$ are the acceleration, velocity and displacement, respectively of the system at time t, and $\ddot{x}_g(t)$ is the ground acceleration at time t. In this simulation, the values of m, c and k used were m=1 kg, c=0.25, k=8. The sampling rate of earthquake acceleration signal is 50 Hz.

The first step is to generate the responses of displacement and acceleration of the structural system under the excitation. The ordinary differential equation (ODE) is solved in MATLAB using 'ode45' function to obtain the time histories of displacement, velocity and acceleration. This displacement taken directly from the model is in effect the true displacement which will be used as the reference to evaluate the working performance of the Kalman filter estimation.

To simulate representative measured data, a white noise is superimposed onto the true acceleration and displacement signals. The noise level with 10% root mean square (RMS) noise-to-signal ratio which is regarded as a normal working range for civil applications is used in this example. The true noise variances of acceleration and displacement measurement are 0.760 m^2/s^4 and 0.0064 m^2, respectively for the discrete time system. The sample rate is taken as 50 Hz for measured acceleration and 10 Hz for measured displacement. The measured signals of displacement and acceleration are shown in Fig. 5.1. The acceleration signal is the upper plot with respect to the y-axis on the left. The displacement signal is the lower plot with respect to the y axis on the right.

5.3.2 Parameter estimation by MLE

This section demonstrates how MLE can be used to estimate the unknown noise parameters which will later be used in the Kalman filter estimation.

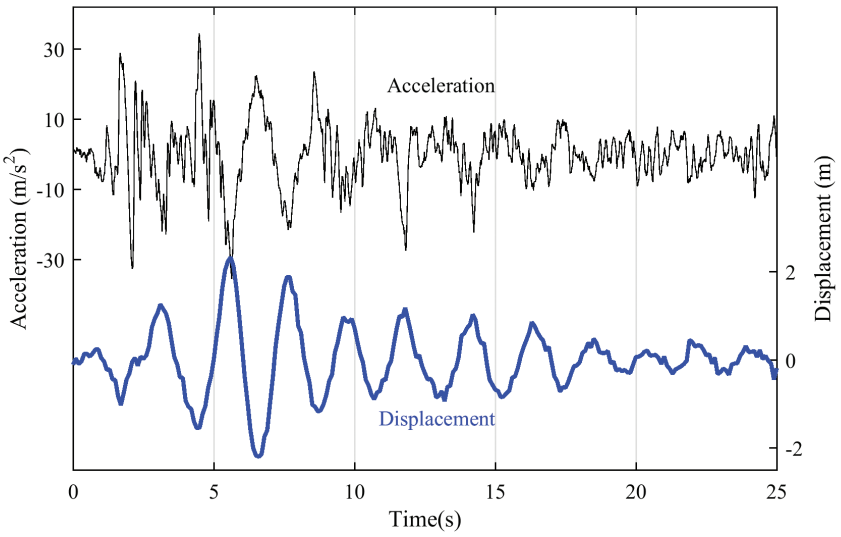


Fig. 5.1 Measured signals of acceleration and displacement of a simulated SOF system under earthquake excitation.

Based on the principles in section 5.2.2, MLE is used to find the unknown parameters θ which maximise the log-likelihood function in Equation 5.9. Given an initial guess θ^0 about the unknown parameters, the optimal parameters $\theta^\star$ are found through the iterations leading to convergence of the log-likelihood function using an optimisation algorithm (Sims, 1999). Instead of using a single initial guess, ten sets of initial guesses $\theta_1^0, \theta_2^0, ..., \theta_1 0^0$ are specified for optimisation separately to derive ten sets of updated parameters $\theta_1^\star, \theta_2^\star, ..., \theta_1 0^\star$ by the MLE method. The updated parameters corresponding to the maximum log-likelihood after optimisation are taken as the global optimum $\theta_j^\star$.

The unknown model parameters $\theta_j^\star = [q, r, \mathbf{X}(0)]$ include the noise variances q and r corresponding to the acceleration and displacement measurement in continuous-time system as well as the initial states $\mathbf{X}(0)$ of displacement and acceleration. The initial states are specified as zero in all ten sets while the initial values of q and r are generated by the random selection between the specified lower and upper bounds. The lower bound for q is selected to be close to zero e.g. $1 \times 10^{-7} m^2/s^3$ while the upper bound is based on an assumed signal-to-noise ratio (SNR). For example, the variance of acceleration signal shown in Fig. 5.1 is calculated to be 79.31 m^2/s^4. Based on the conservative assumption that the SNR value is equal to 0.5, the variance of noise in acceleration signal has an upper bound of 39.66 m^2/s^4. Since the parameter q is the noise variance in continuous-time system, the upper bound of q is derived by dividing the sample rate of acceleration (50 Hz) reaching 0.79 m^2/s^3. Similarly, the lower and upper bounds for the displacement signal are specified as $1 \times 10^{-7} m^2 s$ and 0.011 $m^2 s$, respectively.

The initial guesses (q_j^0 and r_j^0) are indicated in the second and third columns of Table 5.1 with units of m^2/s^3 and $m^2 s$, respectively while the updated parameters ($q_j^\star$, $r_j^\star$, initial displacement and initial velocity) are indicated in columns 4 to 7 respectively. Column 9 (Status) indicates if the algorithm converged or not while Column 8 is the updated log-likelihood. It can been seen that set Nos. 2, 6 and 9 failed to converge which was primarily due to problems with the selected initial values. However, set Nos. 1, 3, 4, 5, 7, 8 and 10 all converged to the same parameter values and log-likelihood value. These values are regarded as the global optimum and will be used for data fusion. The noise variances of acceleration and displacement measurement in discrete-time system deduced by the optimal values $q^\star$ and $r^\star$ (of the continuous-time system) multiplying the acceleration's sample rate (i.e. 0.016 m^2/s^3 $\times$ 50 Hz = 0.080 m^2/s^4) are predicted as 0.80 $m^2/s^4$and 0.0072 m^2, respectively. The estimation of measurement noises by MLE is very close to the true values (0.760 m^2/s^4 and 0.0064 m^2).

Table 5.1 Updated unknown model parameters and corresponding log-likelihood by MLE

#	Initial values θ_j^0		Updated values $\theta_j^\star$ found by MLE					
	q_j^0 (m^2/s^3)	r_j^0 $(m^2 \cdot s)$	$q_j^\star$ (m^2/s^3)	$r_j^\star$ $(m^2 \cdot s)$	Disp(0) (m)	Velo(0) (m/s)	Log-likelihood	Status
1	0.10	0.44	0.016	1.44e-4	-0.064	0.067	-10.19	OK
2	5.34e-3	9.63e-7	5.34e-3	9.63e-7	0	0	-1.39e4	Failed
3	1.01	1.14	0.016	1.44e-4	-0.064	0.067	-10.19	OK
4	1.003	5.00e-4	0.016	1.44e-4	-0.064	0.067	-10.19	OK
5	1.79e-4	0.51	0.016	1.44e-4	-0.064	0.067	-10.19	OK
6	7.78e-3	3.56e-7	7.78e-3	3.56e-7	0	0	-3.49e4	Failed
7	0.011	0.040	0.016	1.44e-4	-0.064	0.067	-10.19	OK
8	7.75e-3	3.16e-6	0.016	1.44e-4	-0.064	0.067	-10.19	OK
9	1.74e-5	4.21e-7	1.74e-5	4.21e-7	0	0	-4.54e4	Failed
10	0.015	3.32e-5	0.016	1.44e-4	-0.064	0.067	-10.19	OK

5.3.3 Estimation of displacement signals by Kalman filter

This section describes how the multi-rate Kalman filter can be used for estimating displacement. To evaluate the performance of sole Kalman filter estimation versus Kalman filter estimation supplemented by the MLE, several sets of noise parameters including the optimal parameters predicted by the MLE are used for displacement estimation separately. Subsequently, the results obtained are compared to the true response, which allowed the accuracy of sole Kalman filter estimation, and Kalman filter estimation supplemented with MLE to be evaluated.

Based on the mathematical principles mentioned in section 5.2.1, the data fusion process involves estimating the mean and error covariance of displacement and velocity through forward Kalman filter and backward smoothing. The estimation could be conducted given the measurement signals of acceleration and displacement, as well as the noise parameters e.g. the noise variances of measured acceleration and displacement. If the sole Kalman filter is applied for the data fusion, the users need to specify the values of measurement noise variances (q and r) according to their judgement or experience that might deviate from the true values. Although noise variances ($q^\star$ and $r^\star$) have been predicted by the MLE in section 5.3.2, a set of different noise variance values is taken for data fusion to study the effect of parameter selection on the estimation accuracy.

To give the reader a feeling for how far the chosen values for noise variance q and r are from the MLE predicted values ($q^\star$ and $r^\star$), we define the variables η_q and η_r which are the ratios between the selected noise parameters and the prediction by MLE, (i.e. $\eta_q = q/q^\star$ and $\eta_r = r/r^\star$). As mentioned in Yuen et al. (2007), the data fusion results (the mean of state variables) remain identical when the values of noise variances (q and r) are scaled by the same factor, therefore the possibility of parameter compensation exists for the mean estimation **X**. For example if values of q and r far from the optimal values of $q^\star$ and $r^\star$ are selected such that η_q=100, and η_r=100, the estimated mean of displacement is the same as that from the case when η_q=1, and η_r=1. Similarly this means that the case of $\eta_q = 1, \eta_r = 20$ will give the same estimation results about displacement as the case of $\eta_q = 1/20, \eta_r = 1$.

Due to the above, when investigating the effect of noise parameter selection on the accuracy of the predicted displacement it makes sense to maintain one of the parameters at its optimum value and to vary the other. In this study the ratio η_q remains unchanged (equal to 1) while the ratio is varied. Five sets of different noise parameters are chosen, shown in Table 5.2. The first set is equal to the optimal values determined by MLE ($\eta_q = 1, \eta_r = 1$) and the other four sets correspond to the overestimation or underestimation of noise variance in displacement measurement. These five sets of variables are used for Kalman filter estimation separately. The estimation results are compared with the true signals from the simulation to study the estimation accuracy.

Fig. 5.2 indicates the error of four separate displacement signals labelled 'Measurement', 'Estimation (MLE)', Estimation (η_r=20)' and Estimation (η_r=100).

- 'Measurement' is the true displacement signal contaminated with white noise (10% RMS noise-to-signal ratio),
- 'Estimation (MLE)' is the estimated displacement signal after data fusion of measured displacement and acceleration when the values of q and r are

Table 5.2 NRMS errors of estimated displacements and velocities by Kalman filter using five sets of noise parameters (q and r)

Set No.	$\eta_q = q/q^\star$	$\eta_r = r/r^\star$	NRMSE of Disp	NRMSE of Velo
1	1	1	0.50%	0.70%
2	1	20	0.81%	0.82%
3	1	100	1.41%	0.88%
4	1	1/20	0.50%	0.73%
5	1	1/100	0.64%	1.05%
Measurement			1.74%	–

predicted by MLE (η_q=1 and η_r=1, essentially this signal corresponds to row 1 of Table 5.2),

- 'Estimation (η_r=20)' is the estimated displacement signal after data fusion when the values of q and r are defined as η_q=1 and η_r= 20 (this signal corresponds to row 2 of Table 5.2), and
- 'Estimation (η_r=100)' corresponds to row 3 of Table 5.2.

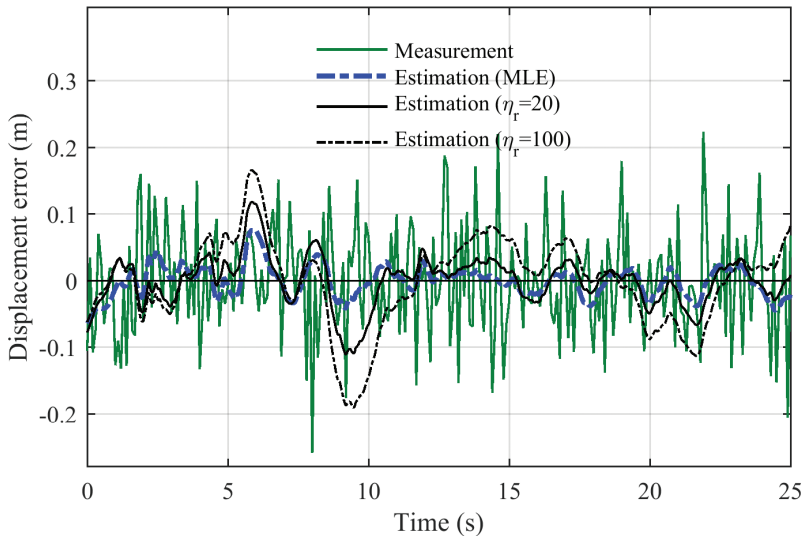


Fig. 5.2 Errors in four displacement signals by comparing with the true one: (i) the measured signal that is the true one superimposed with a white noise (10% RMS noise-to-signal ratio); (ii) the estimated signal by Kalman filter using the optimal estimate of noise parameters ($q^\star$ and $r^\star$) by MLE; (iii) the estimated signal by Kalman filter using the noise parameters ($q = q^\star$ and $r = r^\star \times 20$); and (iv) the estimated signal by Kalman filter using the noise parameters ($q = q^\star$ and $r = r^\star \times 100$).

It can be seen in the figure that the largest errors occur in the measurement signal whereas ('Estimation (MLE))' signal shows the least errors. 'Estimation (η_r=100)' has less high frequency noise than the 'Measurement' signal, however in

a number of areas the amplitude of the errors are of a similar order of magnitude to the errors in the measurement signal. This shows that (i) data fusion brings about improved estimations of displacement; and (ii) using MLE to predict noise variances q and r is likely to result in more accurate estimation of displacement than that by sole Kalman filter where q and r are selected based on judgement/experience. Inaccurate selection of q and r could even lead to displacement estimates that make no accuracy improvement compared with the direct measured signals.

In an effort to further quantify the accuracy of the estimated signals, error analysis is conducted for all the five data sets of Table 5.2, which is carried out using the normalized root mean square error (NRMSE) defined by

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2)}}{max(x_i) - min(x_i)} \tag{5.12}$$

Where $\hat{x}_i, x_i$ denote the estimated and true value at time step i, and n denotes the number of sampled data points.

The NRMS errors for displacement signals 'Estimation (MLE)', 'Estimation (η_r=20)' and 'Estimation (η_r=100)' in Fig. 5.2(b) are 0.5%, 0.81%and 1.4%, respectively. Although not plotted in Fig. 5.2 simulations for situations where the noise is underestimated i.e. η_r=1/20, and η_r=1/100 have been carried out and the associated NRSME values are shown in row 4 and 5 of Table 5.2.

The results in Table 5.2 and Fig. 5.2 show that the selection of the noise parameters has direct influence on the estimation accuracy of displacement data. Without the step of noise parameter (q and r) determination (i.e. applying MLE), the estimation accuracy will be highly dependent on the users' selection of parameters. In order to ensure good performance of Kalman filter, MLE is a necessary step to find the accurate prediction of noise parameters. Through this simulation example, the MLE method is validated to be effective for the estimation of noise parameters and beneficial to the estimation performance of Kalman filter. In the next section the same approach is applied to field data.

5.4 Field test on Humber Bridge

Having validated the proposed approach on simulated data, the next step was to implement it on a challenging full-scale structure, in this case the Humber Bridge in UK.

The purpose of this section is to examine the performance of the proposed data fusion method when applied to field data. GPS is the common choice for deformation measurement in long-span bridges (Brownjohn et al., 2015) however, as discussed in section 5.1, GPS performance with respect to accuracy has some shortcomings. Nickitopoulou et al. (2006) suggest standard accuracy of 35 mm

for vertical measurements at 98.5 percentile level, without gross errors such as cycle slip or multipath. A previous short-term monitoring campaign in the Humber Bridge (Ashkenazi and Roberts, 1997) provided a figure of 3 mm resolution in the vertical direction while a more recent test demonstrates a poorer resolution (Brownjohn et al., 2015). Therefore combining GPS signals with accelerometer measurement using data fusion method should be a good approach to overcome the shortcomings of GPS. During the monitoring campaign at the Humber Bridge, the mid-span displacement in vertical direction was estimated by combining the GPS and accelerometer measurement using the multi-rate Kalman filter and MLE. It is noted that the accelerometers Honeywell QA750 used in this application are DC-response devices that enable capturing low frequency motions. The sensor is reported to have a resolution or threshold better than 1 μg while the studies by Brownjohn and Botfield (2009) indicated broadband noise floor (0-100 Hz) at 3 $\mu m/s^2\sqrt{Hz}$ or better. Piezoelectric accelerometers with no DC response would not be suitable for this method.

In section 5.3.3 the true displacement from the numerical model provided the reference for comparing the displacement signals estimated from data fusion. In the field, a reference displacement is required but not readily available. In this study a vision-based system provided a reference as it was felt that this gave the best performance with respect to scanning frequency and measurement accuracy. Measurement accuracy of vision-based system depends on several parameters, e.g. camera-to-target distance, estimation of camera intrinsic parameters, dimension information and dispersion of target tracking results in image, etc. Khuc and Catbas (2017b) investigated the measurement accuracy of vision-based system in a short-range distance (< 14 m) and suggested an accuracy of 0.04 mm. Martins et al. (2015) demonstrated the uncertainty evaluation of displacement measurement by vision-based system in a long-span bridge monitoring test and illustrated a standard measurement accuracy of 1.7 mm in the vertical direction. In vision-based system, the nominal resolution of target tracking can be better than 0.01 pixel while the reported accuracy varies from 0.5 to 0.01 pixel (Bing et al., 2006) which is related to target pattern (texture contrast) (Busca et al., 2014) and illumination condition. In this application, given the focal length of 300 mm, the camera sensor resolution at 0.0055 mm/pixel and the camera-to-target distance at approx. 710 m, the accuracy of 0.1 pixel (artificial target of high-contrast pattern) in image plane corresponds to an accuracy (or rather resolution) of 1.3 mm in the structural system. However, the vision-based system meets several challenges on site e.g. target pattern change due to lighting condition, camera shake caused by human or wind and partly obstructed view. On the day of the test, environmental conditions were optimal for camera monitoring (calm day, stable light conditions) and from previous experience of using the system in these conditions it is believed

to reflect true displacement with good accuracy. In the text which follows, this section is broken into two parts, section 5.4.1 reports the performance of the proposed approach when applied to field data and section 5.4.2 gives a brief overview on the challenge of varying measurement noise in the displacement signal returned by GPS measurements.

The test has been reported in Chapter 3.2.1 to evaluate the working performance of a commercial vision-based system Imetrum. The test configuration could refer to Fig. 3.2.

5.4.1 Examining performance of data fusion approach

This section reports the results of applying the proposed MLE enhanced data fusion method to the field data recorded on Humber Bridge and for convenience it is broken into three subsections. Section 5.4.1.1 describes the input signals and the displacement signal estimated following MLE and data fusion. Section 5.4.1.2 briefly demonstrates the 'extra' information on the dynamic response of the structure contained in the estimated displacement signal. Finally section 5.4.1.3 compares GPS originating displacement signals (raw and estimated) to the benchmark displacement signal measured by the vision-based system.

5.4.1.1 Applying MLE and Kalman filter to field data

In this section, the data fusion method proposed in section 5.2 was adopted to merge the vertical displacement from the GPS sensor with the collocated acceleration measured by QA accelerometer.

The quasi-static displacement in vertical direction induced by traffic loads is the main interest of this study. Ten-minute signals of vertical displacement by GPS, vertical displacement by vision-based system and vertical acceleration from QA accelerometer are truncated shown in Fig. 5.3, where acceleration and displacement are plotted with respect to the left and right y-axes, respectively. It can be seen in the figure that broadly speaking the GPS displacement agrees very well with the vision-based displacement, this is evidenced by the fact that cross correlation between two displacement signals is 98.4%. For the scale used in this plot it is difficult to accurately quantify the magnitude of the differences between the two signals but in some places it looks to be in the region of 10-20 mm. Further information on the differences between the two signals is provided later in Fig. 5.7.

The first step of analysis is to identify the noise parameters in GPS and accelerometer measurement using the MLE. The noise in GPS and QA accelerometer measurement is assumed to be a white noise process with zero mean and fixed variance and then the MLE proposed in section 5.2.2 is used to estimate the noise variances of these two signals. The estimation process (e.g. selecting the initial

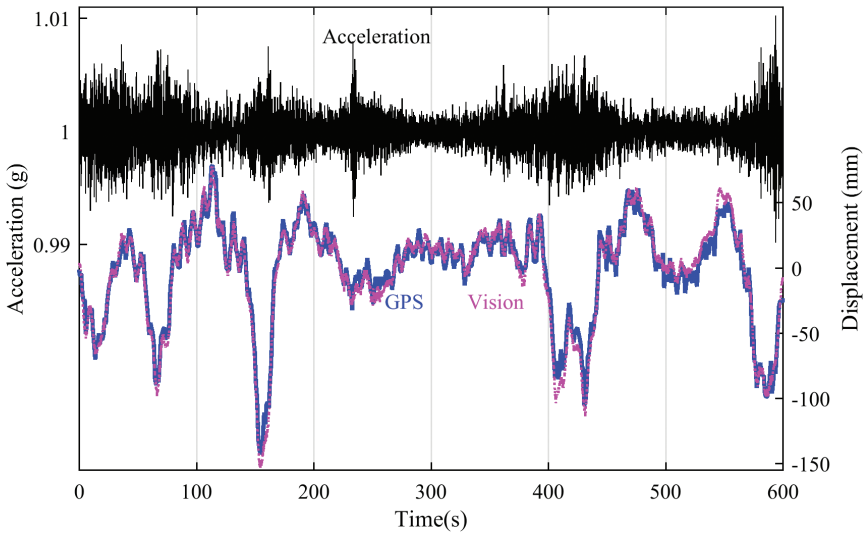


Fig. 5.3 Measured signals by accelerometer, GPS and vision-based system in vertical direction in the Humber Bridge

values for unknown parameters, updating the values for unknown parameters based on an optimisation algorithm and then finding the global optimum) was similar to that for the numerical example presented in section 5.3.2and only the estimation results are demonstrated here. The results of the MLE provide estimates of standard deviations of GPS noise and accelerometer noise in vertical direction as 6.79 mm and 1.77 mm/s^2, respectively (for the signals shown in Fig. 5.3). These results will be adopted in the Kalman filter estimation in the next step.

The second step of analysis is to fuse the GPS and accelerometer measurement in vertical direction by multi-rate Kalman filter algorithm. Prior to fusion, the acceleration is scaled from g (see Fig. 5.3) to mm/s^2. The variation between GPS measurement and down-sampled estimation is regarded as GPS noise and is indicated in Fig. 5.4 with the RMS value of 6.45 mm.

In the following sections (5.4.1.2 and 5.4.1.3), the estimation results by Kalman filter and MLE will be evaluated in frequency domain compared with the measurement by accelerometers and in time domain compared with the measurement by vision-based system.

5.4.1.2 Applying MLE and Kalman filter to field data

The power spectral densities (PSD) of the raw acceleration data, the raw GPS displacement signal and the estimated displacement (following data fusion) were obtained by Welch's method (Welch, 1967). The data were split into 10 equal-length segments with 50% overlap, and each segment was windowed with a Hamming window and the results are shown in Fig. 5.5. From the previous studies on the bridge it is known that (vertical) modal frequencies exist at 0.117 Hz, 0.31 Hz and 0.46 Hz (Rahbari et al., 2015). The GPS system on Humber Bridge has

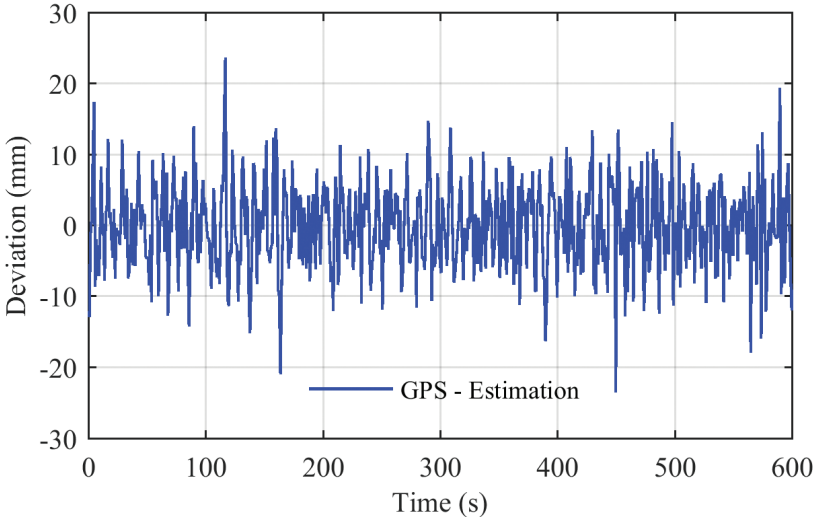


Fig. 5.4 Difference between two displacement time history signals: GPS direct measurement and Kalman filter estimation using the optimal estimate of noise parameters by MLE.

a sample rate of 1 Hz (Brownjohn et al., 2015), which means the sample rate is sufficiently high to capture modal frequencies in the range 0-0.5 Hz. If the GPS plot in Fig. 5.3 is examined, it can be seen that the first vertical modal frequency at 0.117 Hz is identified by GPS measurement, however two other modal frequencies which occur in the 0-0.5 Hz range (namely 0.31 Hz and 0.46 Hz) are not detected. In theory, the GPS measurement has the chance to capture the second and third vertical modes but in fact it has failed to do so. Similarly, the measurement by vision-based system with the sample rate at 10 Hz identifies the first and second modal frequencies (0.117 Hz and 0.31 Hz) and fails to capture more.

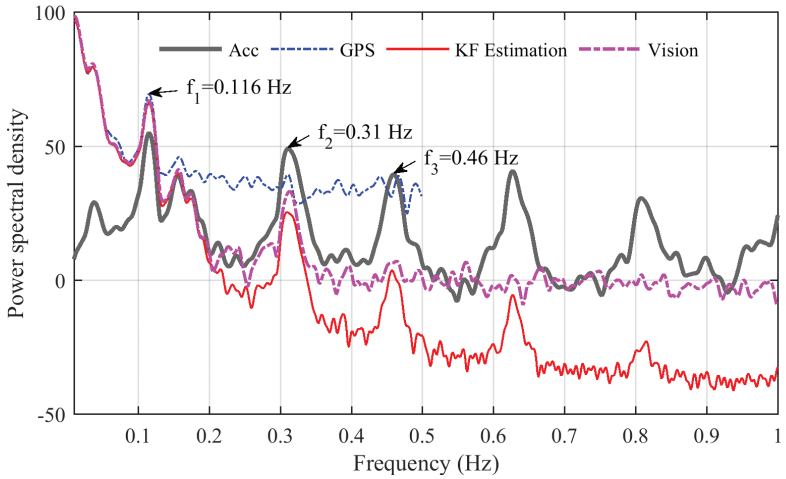


Fig. 5.5 Power spectral densities of four time history signals by Welch's method: (i) the raw acceleration signal; (ii) the raw GPS displacement signal; and (iii) the estimated displacement by fusing the acceleration and displacement measurement using Kalman filter; and (iv) displacement signal measured by vision-based system.

Although both GPS and vision-based system have been validated to be feasible for structural dynamic monitoring (Khuc and Catbas, 2017b; Li et al., 2006), it is difficult in this case to use the displacement sensing system to measure the structural dynamics by providing a high sample rate. This is mainly because the displacement induced by vehicle loads is always dominated by the static and quasi-static components while the dynamic component of displacement is relatively small (i.e. the root mean square of de-trended acceleration signal during this time interval is only 0.016 m/s^2) and easily contaminated by the measurement noise.

In the estimated displacement fused by GPS and accelerometer measurement, the modal frequencies with the value 0.31 Hz and 0.46 Hz are clearly identified in Fig. 5.5. This indicates that the noise in estimated displacement is reduced making the dynamic components clear. The sampling rate of the estimated displacement signal is 20 Hz and so, it also proved possible to identify the modal frequency at 0.63 Hz.

5.4.1.3 Enhanced dynamic information in displacement estimate following data fusion

The measurement by the vision-based system (Imetrum) was adopted to evaluate the estimation results. Before deploying the Imetrum system on Humber Bridge, the system accuracy and stability have been validated through field tests on short-span bridges in Exeter by comparing with the measurement obtained by contact sensors (i.e. LVDT) (Hester et al., 2017). Therefore the measurement by vision-based system is believed to a suitable benchmark to compare with GPS measurement and estimation by data fusion.

If Fig. 3.2(b) is examined it can be seen that the artificial target, GPS rover, and vertical accelerometer were not at the same transverse position on the deck. The artificial target is mounted on the parapet, the GPS rover is mounted on the main cable 3 m from the parapet and the accelerometer is on the inside of the box girder a further 2.5 m from the parapet. Therefore it is necessary to compensate for the influence of torsional deck rotation/tilt before any comparison. Since the deck rotation under traffic loads is dominated by quasi-static components with very small dynamics, the influence of location deviations on vertical acceleration measurement could be ignored while the influence on vertical displacement measurement should be considered. The GPS rover was mounted at the top of main cable, 11 m away from the centre of box section. The artificial target was located at the edge of walkway cantilevers, about 14.5 m away from the centre. Only the quasi-static tilt of the deck was considered and corrected which was estimated by applying low-pass filter to the height difference of two GPS rovers located at mid-span (one rover on each of the two main cables). The filtered result is validated by cross checking it with the rotation angle of deck predicted from

low frequency component of lateral acceleration shown in Fig. 5.6 which senses rotation via the component of gravity resolved along the DC accelerometer axis. The moving trends of two predictions look similar and the GPS height variation after low-pass filter was adopted for tilt correction. Assuming that the deck rotates around the centre as a rigid body, the tilt was calculated by the rotation angle and the distance to deck centre. The tilt difference between vision-based system and GPS measurement is then removed from the raw measurement.

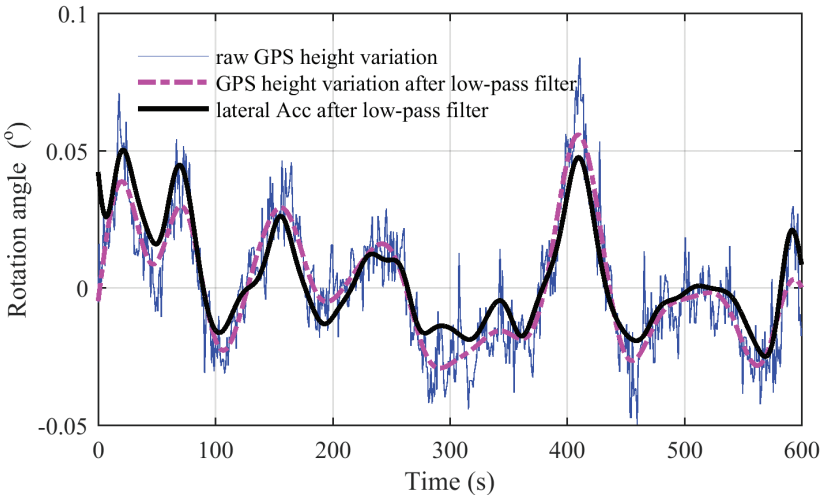


Fig. 5.6 Predicted rotation angle of deck in the longitude direction at mid-span by: (i) variation of the two GPS height measurement at two sides of mid-span; (ii) low-pass filter results of variation of the two GPS height measurement (cut-off frequency at 0.02 Hz); and (iii) DC component of lateral acceleration measurement at mid-span by low-pass filter (cut-off frequency at 0.04 Hz).

In this section three different displacement signals originating from GPS data are compared to the benchmark vision-based displacement signal: the raw GPS measurement (GPS), and Kalman filter estimation using two different values of η_r (Estimation η_r= 1, and Estimation η_r= 200). Here η_q and η_r have the same meaning as in section 5.3.3, and for both estimated displacement signals η_q= 1. The purpose of using two different values of η_r is to demonstrate the difference in the estimated displacement when noise parameters are well or poorly selected. The results of these three displacement signals subtracted by the optical measurement (down-sampled to equal sample rate as GPS measurement) are shown in Fig. 5.7.

It can be seen that the largest errors occur in the 'GPS' measurement signal, Estimation (η_r=1) shows the least errors, and Estimation (η_r=200) has errors somewhere in between. (Note, 'GPS' data points in the figure are in effect the differences between the two displacement signals shown earlier in Fig. 5.3 after the difference in the transverse position of the two sensors is corrected for). The errors evident in Fig. 5.7 can be summarised as normalised root mean square error (NRMSE) values as defined in Equation 5.12. The NRMS error of displacement signal for GPS measurement, Estimation (η_r=1) and Estimation (η_r=200) are 3.17%,

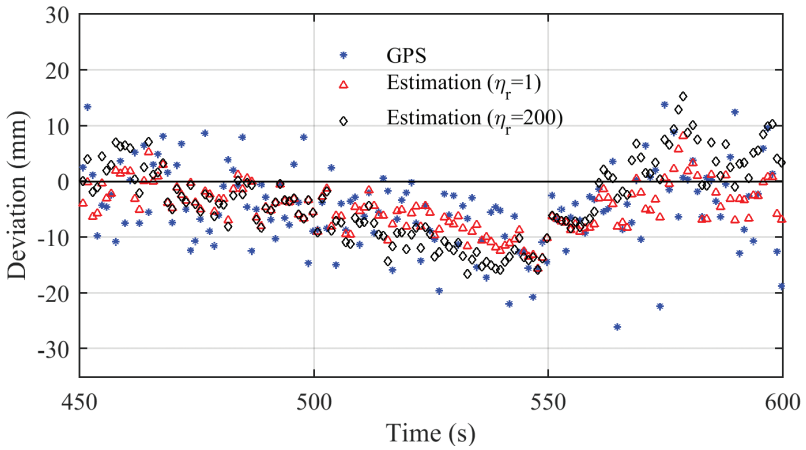


Fig. 5.7 Comparison of three displacement signals with the measurement by vision-based system: (i) GPS raw measurement; (ii) Kalman filter estimation using the optimal noise parameters ($q^\star$ and $r^\star$) predicted by MLE; and (iii) Kalman filter estimation using the noise parameters ($q = q^\star$ and $r = r^\star \times 200$).

2.37% and 2.90%, respectively. The results show that using data fusion results in displacement signals that are more accurate than the raw GPS displacement signal with the NRMS error decreased from 3.17% to 2.37%, and moreover that using MLE to identify the optimal estimate of noise parameters results in a more accurate displacement estimation than that when noise parameters are specified based on subjective judgement.

As a final comment on Fig. 5.7, it is evident that there is some low-frequency variation between the GPS originating signals and the vision-based measurements. This might be caused by the difference of sensor locations, the measurements from two sensors are not perfectly correlated to the deck movement i.e. there might be some local movements in the main cable where GPS rover was mounted or in the parapet where the artificial target was attached.

5.4.2 Comparing raw GPS and estimated displacements to benchmark vision-based displacement

Having seen in Fig. 5.7 the errors that can result in the estimated displacement signal due to incorrectly estimating the noise present in the raw GPS displacement measurements, this section examines how the GPS noise varies over the course of the day. The fact that the noise in GPS measurements is not constant means that if using GPS measurements in a data fusion system it is necessary to update the values of q and r at certain time intervals using MLE. It is noted that the noise parameter estimation step using MLE is not a real-time processing and instead is to deal with offline data. The text below describes the variation in the noise characteristics of the raw GPS measurement observed on the day of the test at Humber Bridge.

The positioning accuracy of GPS measurement depends mainly on the atmosphere condition, satellite geometry and the possible secondary reflection or multi-path (Nickitopoulou et al., 2006). Therefore the GPS noise should be time dependent even during a single day, since the satellites are not geo-stationary. To investigate the variation in GPS noise over the course of the day, two ten-minute signals of GPS and QA accelerometer measurement in vertical direction were truncated for data fusion. The first time interval was from 12:35 to 12:45 and the second one was from 18:50 to 19:00. The measurement signals after time synchronisation are shown in Fig. 5.8(a) and (b) respectively.

Firstly, the MLE method was applied to estimate the noise parameters of GPS and accelerometer measurement for both time intervals. The result indicates that the standard deviations of GPS noise were estimated as 7.28 mm and 13.70 mm for time intervals 1 and 2 respectively. The GPS measurement in the second time interval (18:50 to 19:00) is noisier than in the first one, i.e. it has a larger standard deviation of noise.

To allow the larger noise of interval two to be visualised, the estimated noise parameters are used for data fusion by the Kalman filter and the resulting estimated displacements subtracted from the raw GPS measurements are shown in Fig. 5.8(c) and (d) as approximate GPS measurement noise. The RMS of estimated noise was 7.10 mm and 13.36 mm, respectively. The signals in Fig. 5.8 show that the GPS signal in the second time interval is much noisier than the first time interval and that the noise variance of GPS measurement varies with time.

Having seen in Fig. 5.8 the time-dependent nature of GPS noise, it deserves attention to deal with the long-time signals using the proposed data fusion method. In the Kalman filter, the GPS noise is assumed as a Gaussian process with zero mean and fixed variance while the value of GPS noise variance is pre-determined by the MLE using offline measurement data. For a robust application of this method over longer time e.g. several hours or days, it is suggested to update the GPS noise parameters every half hour e.g. truncating the first five minutes of 30-min data for noise parameter estimation using the MLE and applying the tuned noise parameters for data fusion of the whole 30-min signals.

Since the number of available satellites have impact on the GPS positioning accuracy, Fig. 5.9(a) provides the time-dependent characteristics of available satellite over the whole day. It is observed that the number of available satellites from 05:00 AM to 11:30 AM decreased to 3 which are not enough for 3D positioning. Therefore the GPS displacement data during this period was not available. Based on the available GPS data, the MLE was applied to estimate the noise parameters during the data fusion of GPS and accelerometer measurement in vertical direction. The estimated results are shown in Fig. 5.9(b) with the standard deviation of GPS noise varying from 6.16mm to 16.19 mm.

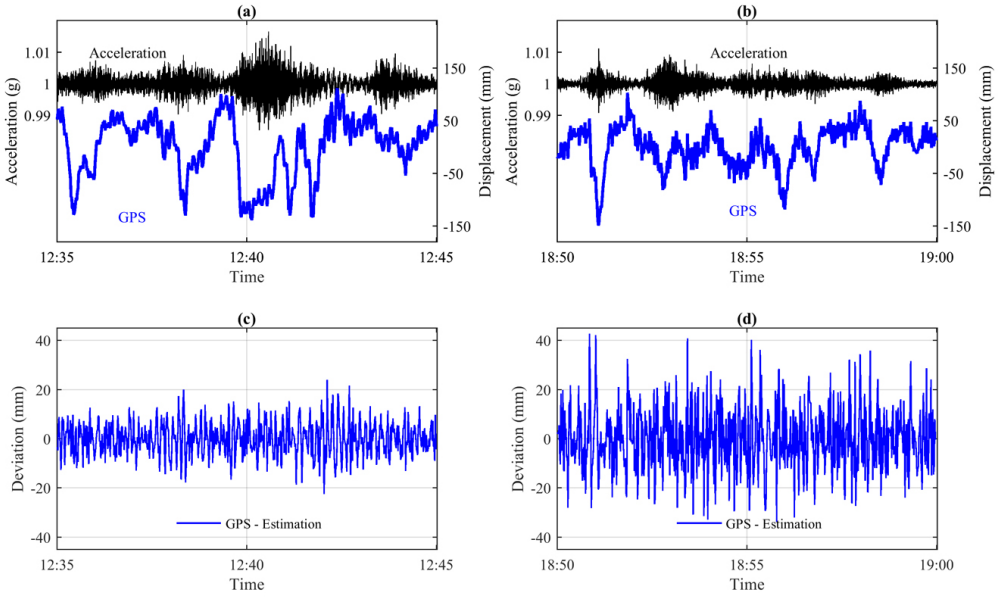


Fig. 5.8 Variation of GPS noise with time: (a) raw acceleration and displacement signals for time interval 1; (b) raw acceleration and displacement signals for time interval 2; (c) approximate GPS noise for time interval 1; and (d) approximate GPS noise for time interval 2.

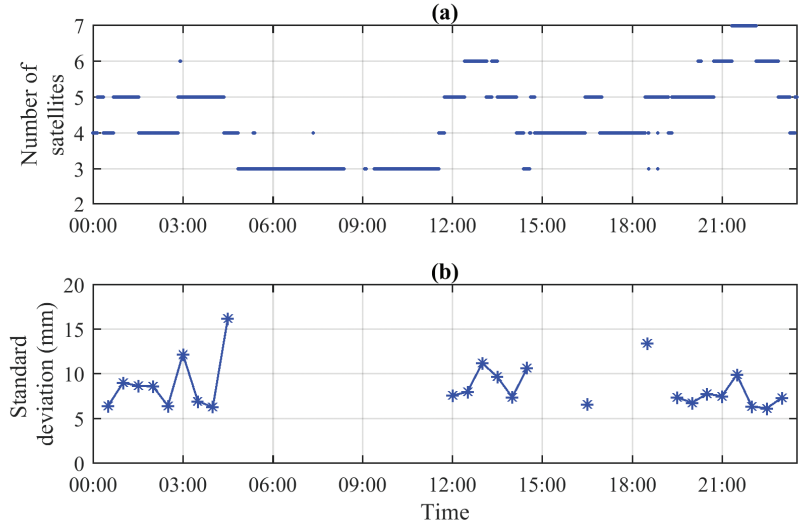


Fig. 5.9 Characteristics of GPS measurement over the whole day: (a) number of available satellites; and (b) estimation results of standard deviation of GPS noise by the MLE.

5.5 Conclusions

Data fusion of displacement and collocated acceleration signals using Kalman filter and maximum likelihood estimation is validated to be effective for improving displacement accuracy. MLE is effective at accurately determining the noise parameters of measurement signals which are required by the application of Kalman filter. Incorporating the MLE could overcome the limitations of parameter

selection according to users' judgement and ensure the good performance of Kalman filter.

The data fusion method is also practical to field implementations. The results in the Humber Bridge test indicate the effectiveness to improve the GPS accuracy and to widen the frequency bandwidth of displacement signal. Compared with the vision-based measurement, the normalised root mean square deviation of GPS measurement was decreased from 3.17% to 2.37% after applying the data fusion. The time-dependent feature of GPS noise is also clearly identified through analysis on the data acquired from different time in the test day with the standard deviation varying from 6 mm to 16 mm.

Finally although the vision-based system is not the focus of this chapter, this study is also a good demonstration about the potential of vision-based system for long-range monitoring.

Chapter 6

Deformation monitoring using a mixed system

Vision-based monitoring faces several field challenges and might fail to achieve desired accuracy level in some real-world test conditions. To make the best use of the potential of vision-based systems, a mixed sensing system consisting of a camera and an accelerometer is proposed for accurate bridge deformation monitoring.

The mixed system implements the custom-developed vision-based system (in Chapter 4) for displacement measurement and considers automatic compensation of camera shake. The data fusion method proposed in Chapter 5 is involved for noise reduction. The system is demonstrated through a field monitoring test on a short-span railway bridge and is validated to offer higher accuracy and wider frequency range than using a camera alone. Displacement data by the mixed system are demonstrated to be viable for estimating bridge influence line, indicating the potential for bridge condition assessment.

In this chapter, measurement by a commercial vision-based system Imetrum was used as a reference to evaluate displacement data acquired directly from GoPro camera and from the mixed system. It is noted that the comparison work is not to validate this commercial system but to demonstrate the working performance improvement of a cheap custom-developed camera system through fusing with collocated accelerometer data.

This chapter is reproduced from a paper titled "**Accurate deformation monitoring on bridge structures using a cost-effective sensing system combined with a camera and accelerometers: case study**", co-authored with James Brownjohn and Farhad Huseynov, which has been accepted to publish in *Journal of Bridge Engineering*.

6.1 Introduction

Vision-based displacement monitoring faces several field challenges and might fail to capture the nature of the structural deformation due to site and environmental conditions e.g. camera and support motion induced by wind or human behaviours (Ribeiro et al., 2014), uncontrolled lighting variations due to cloud passing (Chen et al., 2017), low-contrast target patterns and pattern changes due to obstruction and rain drops (Brownjohn et al., 2017).

Although measurement accuracy of vision-based systems has been validated in some application examples (Ehrhart and Lienhart, 2015c; Feng et al., 2015a; Khuc and Catbas, 2017a), the measurement quality is time-varying and environmentally dependent. Several undesired test conditions (e.g. camera shake, changes of tracking patterns and ambient lighting changes) could possibly reduce the measurement accuracy significantly. It is impossible to avoid all these unsatisfactory circumstances in field testing and these influences are rarely considered in existing studies concerning the development of vision-based systems.

The main purpose of this study is to investigate low-cost options for accurate displacement monitoring on bridge structures. To make the best use of the potential of vision-based systems, a mixed sensing system consisting of a consumer-grade camera and an accelerometer is proposed. Compared with vision-based systems described in existing literature, the displacement output from this mixed system considers the compensation of camera shake automatically and is capable of achieving a higher accuracy level and wider frequency bandwidth even for relatively low-contrast target patterns. Compared with similar work mixing vision-based systems with accelerometers (Chang and Xiao, 2010a; Park et al., 2018), the data fusion method proposed in Chapter 5 is an autonomous implementation without any user supervision or involvement. This mixed system could be implemented for applications where accurate and high-resolution displacement data are required and where the structure can be accessed e.g. for model calibration and estimation of vehicle weights. In this study, the measured displacement is interpreted for estimating bridge influence line effectively by field measurement.

The proposed system is demonstrated through a field monitoring test on a short-span railway bridge during the passing of several trains, leading to a discussion about its working performance. The undesired test conditions considered include apparent camera shake and low-contrast patterns while the evaluation criterion is the accuracy level in the time domain. Although the data fusion method has the capacity to widen the frequency bandwidth of the estimated displacement time histories, capturing bridge dynamics is not the focus here. This is because traffic-induced deformations for road and railway bridges are always dominated by static and quasi-static components, while the dynamic components with relatively low

signal-to-noise ratios are easily contaminated by measurement noise. Therefore, bridge dynamic information is better suited to measurement using accelerometers.

Since direct access to bridge structures is necessary for accelerometer installation in this mixed system, there is a kind of trade-off between having a high accuracy level and the benefit of non-contact sensing. When the signal-to-noise ratio is acceptable, the consumer-grade camera in this mixed system could be used separately for completely non-contact and multi-point displacement (deformation) measurement in bridge applications. The performance of a single consumer-grade camera system is also evaluated in the field monitoring test.

6.2 Methodologies

The vision-based system developed for displacement monitoring in this study consists of a consumer-grade camera (GoPro Hero 4) for video recording and a post-processing package programmed in C++ for video analysis. The main algorithms have been reported in Chapter 4 and the difference in this study is that the influence of camera shake is considered automatically within the calculation process.

A detailed flowchart of the video processing procedure is provided in Fig. 6.1 including three main steps: camera calibration, target tracking and displacement calculation. In the camera calibration step, lens distortion parameters are calibrated in the laboratory ahead of field testing using a chessboard pattern with known dimensions. Projection distortion is often observed on site due to the optical line of sight not being perpendicular to the plane of motion of structural components. To consider the projection distortion, a transformation matrix (i.e. planar homography) is determined, assisted by some control points with known geometric information. Based on the point correspondences between structural coordinates of these control points and image coordinates of their projections, the projection transform is estimated using least-squares optimisation to the total re-projection error. The control points used for calibration could be edge points of pre-installed artificial targets or points from bridge components with known dimensions.

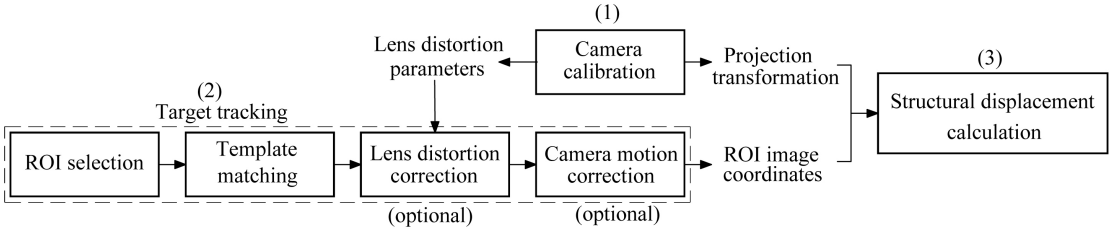


Fig. 6.1 Flowchart of video processing procedures

In the second step, correlation-based template matching is used to localise the ROIs in video frame sequences. To consider lens distortion influence, one feasible

way is to correct video frames before the tracking step, but this is computationally very expensive. In the method used here, the correction occurs not to the full frame but only to the image coordinates estimated from raw frames, saving computational efforts.

When apparent frame shake is observed, a reference ROI around adjacent stationary objects visible in the frame e.g. foundation walls or bridge towers is also tracked. The camera motion is then compensated by subtracting the nominal motion of this reference target. This method has been implemented in several existing studies (Feng and Feng, 2017; Yoneyama and Ueda, 2012) and the difference in this study is that an automatic evaluation process for camera mounting condition is added to determine the necessity of camera motion correction. Two parameters are used to evaluate camera motion occurrence, i.e. the root means square (RMS) and the maximal value of the tracked motions for the stationary target as shown in Fig. 6.2. Tracking accuracy using correlation-based template matching method varies from 0.01 pixel to 0.08 pixel depending on target patterns through an evaluation study in laboratory conditions (reported in Chapter 4) and might become poorer in field testing. An appropriate threshold for the image motion RMS is 0.05 pixel, considering the low-contrast feature of stationary natural targets. The maximum threshold is implemented on time series of the image motion after low-pass filtering to avoid the influence of abnormal events (e.g. sudden partial obstruction on targets) and the threshold is set as 0.08 pixel.

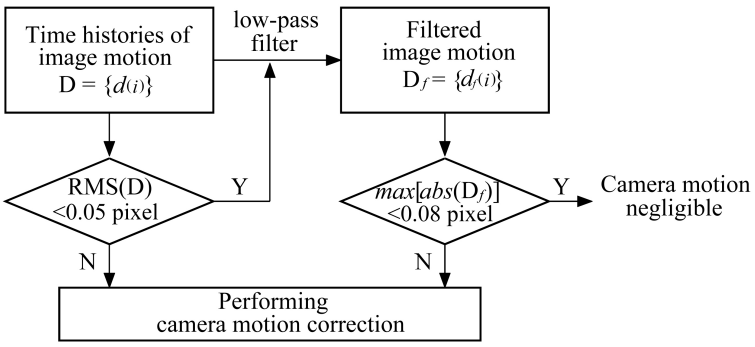


Fig. 6.2 Evaluation criteria for camera stability condition

Finally, the two-dimensional structural displacement is derived based on image coordinates of the ROIs and the projection transform matrix.

Displacements output by the proposed mixed system are the results of the fusion of displacement data from a sole camera and the collocated accelerometer measurement. A field monitoring test on a short-span railway bridge performed for system validation is reported in the next section.

6.3 Field test on a railway bridge

This section describes a field monitoring test on a railway bridge for train-induced displacement measurement.

The Mineral Line Bridge shown in Fig. 6.3 is a skew steel girder bridge with the span length of 14.7 m (from a bridge reconstruction drawing), carrying the West Somerset Railway near Watchet in the UK. A single day of field measurements on 5th September 2017 was used to monitor the deformation induced by passing trains.

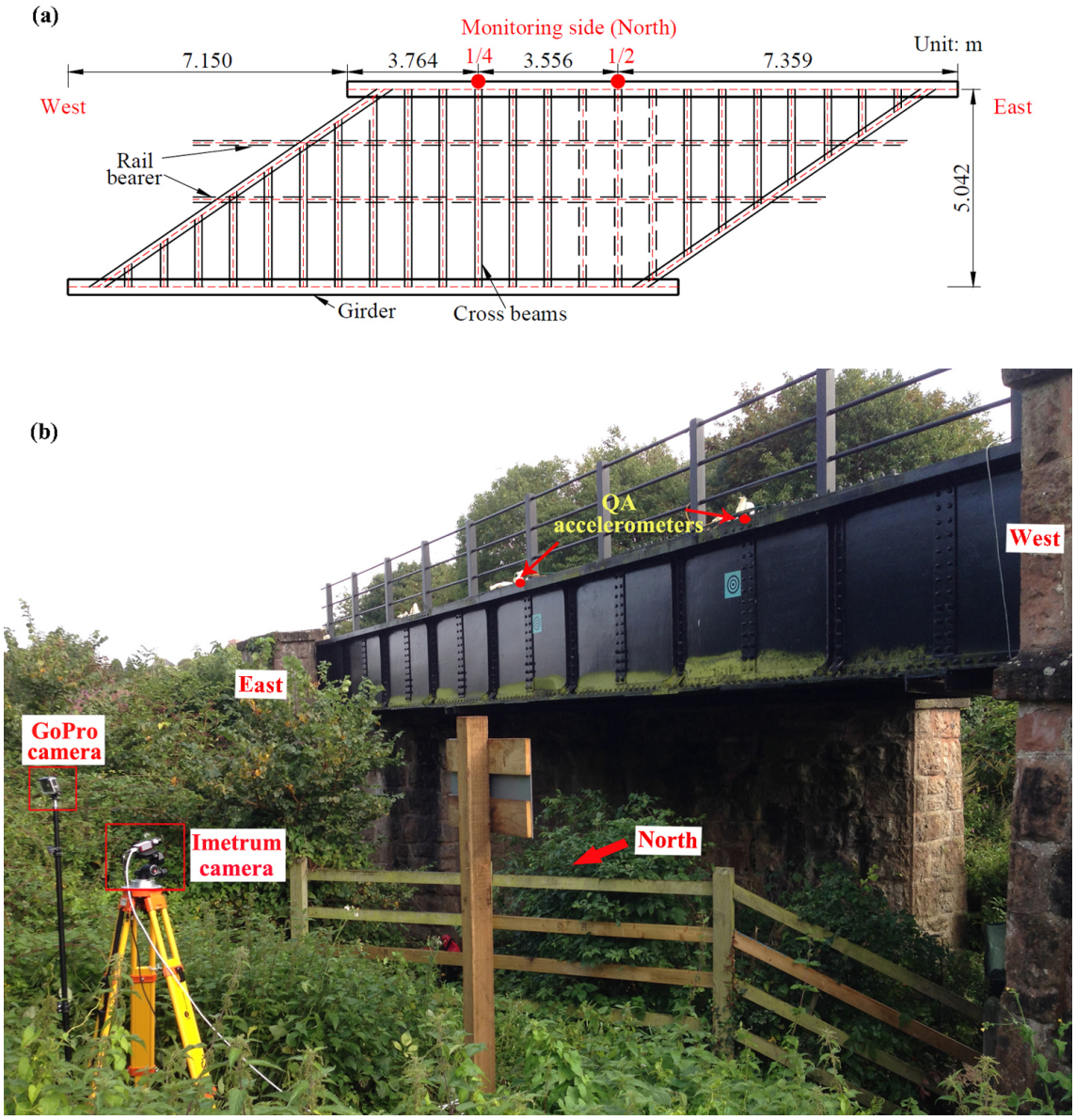


Fig. 6.3 Bridge plan (a) and sensor locations (b)

Three sensing systems were involved in this test indicated in Fig. 6.3(b), including one consumer-grade camera (GoPro Hero 4), a commercial vision-based system (Imetrum dynamic monitoring system, DMS) and two accelerometers, all located in the north side of bridge. The test aimed to evaluate the effectiveness of

two systems (i.e. a consumer-grade camera system and a mixed system combined with a consumer-grade camera and accelerometers) for accurate displacement sensing through comparison with the reference sensor (Imetrum DMS).

GoPro Hero 4, a consumer-grade camera (cost $400) was mounted on a portable tripod stand, 6.9 m from the mid-span of the bridge. The frame rate was set as nominally 24 Hz and the image resolution was 1920 $\times$ 1080 pixels. A narrow field of view setting was selected with the corresponding focal length equivalent to 30-34 mm. The recorded video files initially stored in the camera flash memory card, were copied to a computer for the post-processing to extract displacement time histories.

The Imetrum DMS is a commercial vision-based monitoring system developed by Imetrum Limited, UK and comprises one GigE professional camera and a controller containing a real-time video processing software Video Gauge (VG) for video acquisition and analysis. The Imetrum camera equipped with 50 mm focal length lens was arranged on site adjacent to the GoPro camera. As shown in Fig. 6.3(b), a surveyor's tripod was used for the Imetrum camera, being more stable than the portable tripod stand for the GoPro camera. The frame rate was 30 Hz and the image resolution was 2048 $\times$ 1088 pixels. The auto-exposure feature was switched on in VG software for the automatic adjustment of exposure according to real-time lighting condition to ensure brightness constancy in selected target regions. The Imetrum DMS has been evaluated on several short-span and long-span bridge monitoring tests providing reliable and accurate measurements (Hester et al., 2017; McCormick et al., 2014; Xu et al., 2017), and displacement resolution was found to approach 0.1 mm in a field of view of 20 m (McCormick et al., 2014). Thus, the Imetrum DMS is used as the reference sensor in this study. It is noted that, however, the goal of the proposed mixed system is not to achieve similar performance as this commercial product (Imetrum DMS), but to overcome some general limitations faced by any vision-based system. Stable working performance of the Imetrum DMS was ensured in this study via a high quality tripod and the camera auto-exposure function for brightness constancy in selected target regions. Thus, the Imetrum measurements in this study fortunately avoided the undesired test conditions discussed later for the GoPro measurement, i.e. apparent camera shake and low-contrast patterns. The mixed sensing system is also applicable to improve accuracy of the Imetrum measurements with less than perfect test conditions.

The QA-750 accelerometers are DC-response devices with a resolution better than 1 μg and sensor noise floor of 7 $\mu g/\sqrt{Hz}$ in 0-10 Hz band. The two uni-axial accelerometers located vertically were attached to the bridge's top flange using magnets at approximately mid-span and one-quarter span points, and the sample rate for data acquisition was set to 512 Hz.

The daytime records (lasting seven hours) include the passages of nine trains in total. Considering one train passing the bridge in less than 40 seconds approximately every 50 minutes, monitoring systems took records only around train passages based on the train timetable. The Imetrum system has a remote controller to start/stop video acquisition, thus the camera was not touched during whole recording periods except when adding a waterproof covering to protect against light rain. The GoPro camera was switched on/off by manually pressing one control button thus the camera position could possibly change slightly between different runs. Three runs of measurement data involving passing trains are presented in this study and the information is summarised in Table. 6.1. Trains passing in Run 1 and 2 are of similar type consisting of one steam locomotive and eight carriages. The difference between the two runs is that the GoPro mounting arrangement was stable in Run 1 while apparent camera shake is observed in Run 2. Run 3 corresponds to the records during the passage of a diesel multiple unit train comprising three carriages but no locomotive. The maximum bridge deflection at the mid-span was approximately 3 mm, less than half the maximum deflection in Run 1 and 2, thus requiring higher measurement resolution. GoPro records in Run 3 also include the influence of considerable camera motion.

Table 6.1 Record information about three runs involving train passages

Run #	Train type	No. of carriages	Train speed (km/h)	Maximum deflection	GoPro videos involving camera shake
Run 1	Steam train	8	16	6.87 mm	No
Run 2	Steam train	8	27	6.51 mm	Yes
Run 3	Diesel train	3	34	2.92 mm	Yes

The measurement data in Run 1 are presented first to demonstrate the working performance of a sole camera system in a desired test condition (i.e. stable camera mounting and no observable change of target patterns). The data in Run 2 and 3 are used to validate the effectiveness of improving poor data due to camera motion and low-contrast target patterns through fusion with acceleration data.

6.4 Displacement monitoring using a sole camera

This section demonstrates displacement information extraction from video files recorded using the consumer-grade GoPro camera. The measurement accuracy of a sole camera system through tracking both the artificial and natural targets in Run 1 is evaluated by comparison with the Imetrum DMS reference data.

6.4.1 Video processing process for GoPro records

One sample frame in a GoPro video is indicated in Fig. 6.4(a) that includes apparent image distortion e.g. the parapet railings appear slightly bent. The lens distortion parameters were pre-determined in the laboratory and were used to remove lens distortion influence with the corrected frame in Fig. 6.4(b).

Fig. 6.4 One sample frame by GoPro camera before and after removing lens distortion: (a) before correction; and (b) after correction.

The step of camera calibration also involves estimating projection transformation (i.e. planar homography) through existing dimensions projected in the corrected frame. The geometric information used for calibration is from the width and height of artificial targets (T10 and T20 in Fig. 6.4(a)) attached to the bridge girder, both 200 mm. The planar homography matrices were estimated separately

for the mid-span and one-quarter span targets using least-squares optimisation according to point correspondences.

In the second step of target tracking, a few regions of interest (ROIs) indicated in Fig. 6.4(a) were selected for analysis. The ROIs T10 and T20 are artificial targets with diffuse concentric ring patterns at the mid-span and one-quarter span of the bridge that were tracked by both the GoPro system and the Imetrum DMS. The measurement outputs by the two vision-based systems are compared to evaluate the GoPro system measurement. The ROIs T11 and T21 are natural feature targets including rivet patterns adjacent to the artificial targets. They were analysed in GoPro system to evaluate the feasibility and accuracy level for measuring structural features when direct access to the bridge is not available. The ROI T00 is one natural feature target located at the surface of the stationary masonry wall foundation and was tracked for correcting undesired camera shake when necessary. The ROI locations in the raw frame were estimated using correlation-based template matching algorithm and then transformed to image coordinates in the corrected frame to consider lens distortion influence.

The tracking results for the ROI T10 along the image height direction during a train passage in Run 1 are shown in Fig. 6.5: the left and right axes correspond to the derived locations in the image plane before and after lens distortion correction, respectively. The main difference between two time-history signals is the relative location in the image plane instead of motion amplitudes. This indicates that lens distortion correction step is not essential when using a scale factor to convert the target motion in the image plane to structural displacement. However, this step is necessary when other types of projection transformation (e.g. planar homography or full projection matrix) are implemented since the target location sequences in the image plane are taken for structural displacement calculation.

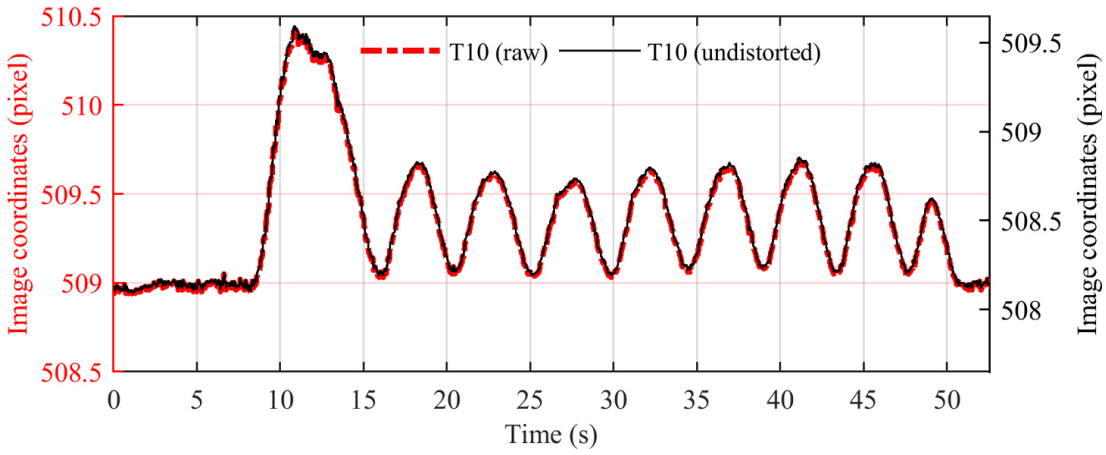


Fig. 6.5 Time series of image coordinates for the target T10 along image height direction before and after removing lens distortion influence

To evaluate the camera mounting stability condition, the stationary region T00 was tracked with the results indicated in Fig. 6.6. The RMS of the image motion (raw) along the image height direction is 0.035 pixel while the maximum deviation after the low-pass filter (with the cut-off frequency of 1 Hz) is 0.056 pixel. Since they are both within the threshold range, the step of camera motion correction is skipped.

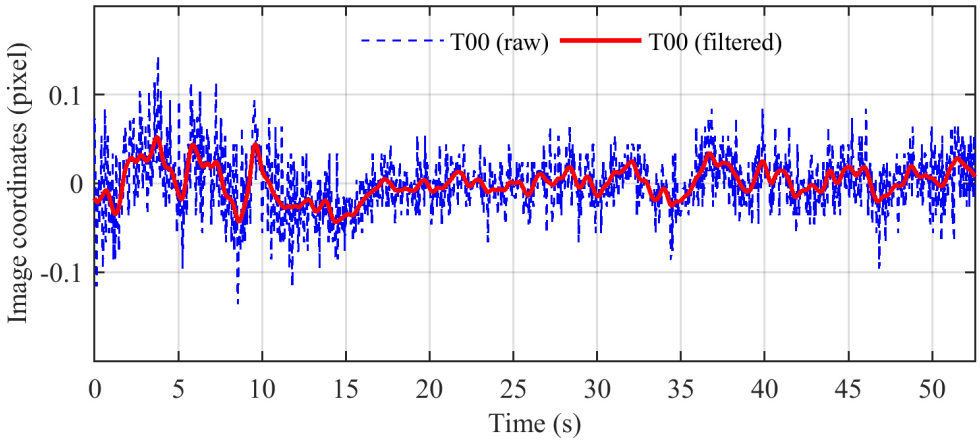


Fig. 6.6 Time series of image motions for the target T00 along image height direction

Finally, the two-dimensional bridge displacement along the vertical and longitudinal directions were calculated based on the output of camera calibration and target tracking steps.

6.4.2 Displacement measurement in Run 1

The train that passed the bridge in Run 1 consisted of a locomotive and a tender followed by eight carriages as shown in Fig. 6.7 (that is taken from the south of the bridge, opposite of the monitoring side). Fig. 6.8 and Fig. 6.9 indicates the measured deformation in the vertical direction at mid-span and one-quarter span during the train passage.

At the mid-span of the bridge (Fig. 6.8), the maximum displacement induced by the locomotive at approximately 6.8 s is measured as 6.87 mm by the Imetrum DMS, 6.77 mm and 6.84 mm by the GoPro system tracking artificial (T10) and natural (T11) targets, respectively. Taking the measurement by the Imetrum DMS as the reference, the cross-correlation coefficients for the GoPro measurement reach 99.8% and 99.4% for tracking the artificial and natural feature targets, respectively. The measurement error for the GoPro measurement is presented in Fig. 6.8(b) through subtracting the reference (after interpolation to the same sample rate) with the root mean squares (RMS) of 0.11 mm and 0.22 mm.

At one-quarter span (Fig. 6.9), the maximum displacement measurement during the locomotive passage is 4.83 mm by the Imetrum DMS, 4.90 mm and 4.77 mm

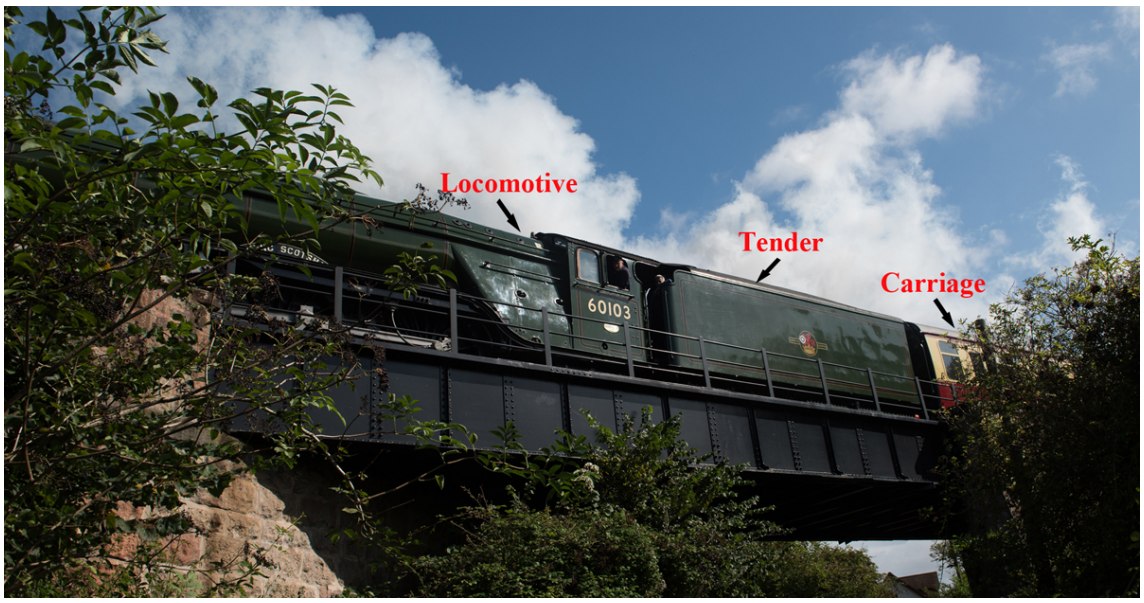


Fig. 6.7 Run 1: a steam train passing the bridge

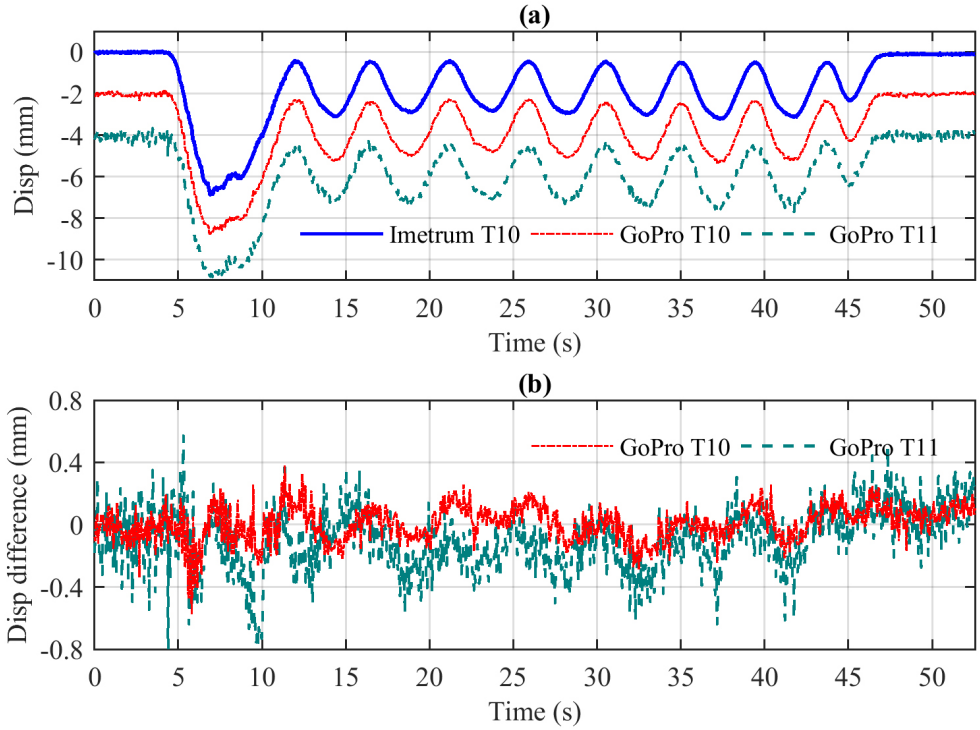


Fig. 6.8 Measured displacement in the vertical direction at bridge mid-span by two vision-based systems: (a) displacement measurement (curves are shifted along y axis for clarification); and (b) the GoPro measurement error evaluated by the reference Imetrum DMS

by the GoPro system for artificial (T20) and natural (T21) targets, respectively. The cross-correlation coefficients between the GoPro measurement and the reference (by the Imetrum DMS) are 99.9% and 99.5%, respectively for tracking artificial target and natural feature patterns. Evaluated against the reference, the RMS of measurement difference using the GoPro system is quantified as 0.09 mm and

0.11 mm. This is slightly reduced compared with that at mid-span probably due to the decreased distance to the camera.

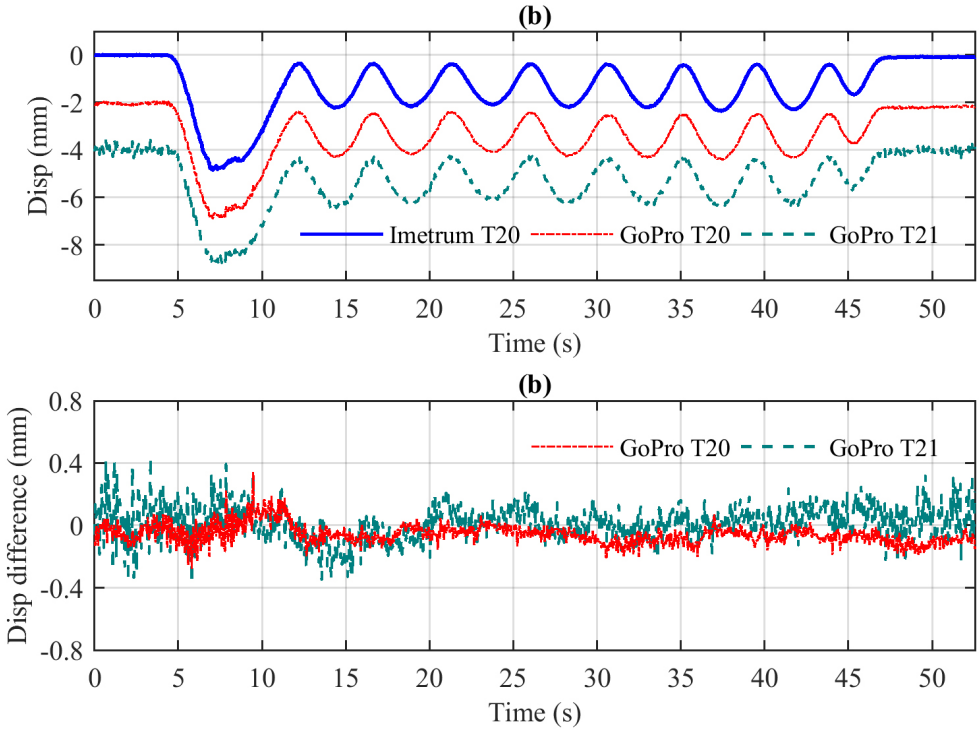


Fig. 6.9 Measured displacement in vertical direction at bridge one-quarter span by two vision-based systems: (a) displacement measurement (curves are shifted along y axis for clarification); and (b) GoPro measurement error evaluated by the reference Imetrum DMS

Measurement noise during the stationary periods (including the first four seconds and the last five seconds) is evaluated using the root mean square (RMS) of measured data and the results are indicated in Table. 6.2. Compared with the Imetrum DMS, the measurement results using the GoPro camera system include a larger noise level when tracking the same artificial targets (T10 and T20). The noise range could be more than doubled (with the RMS reaching 0.16 mm) when the tracked targets are less distinctive e.g. using the natural features (T11 and T21). This phenomenon is accordant with the application preference of high-contrast patterns in digital image correlation field (Schreier et al., 2009). The measurement noise at mid-span is slightly larger than that at one-quarter span possibly due to the increased camera-to-target distance.

Table 6.2 Evaluation of measurement noise during the stationary periods in Run 1

RMS of noise (mm)	Artificial targets		Natural patterns	
	T10	T20	T11	T21
Imetrum DMS	0.02	0.01	–	–
GoPro	0.07	0.04	0.16	0.14

Results indicate that the GoPro system alone could provide accurate measurement of train-induced bridge displacement. If direct access to the bridge is not allowed, the system is capable of performing non-contact displacement monitoring through tracking existing natural patterns with the noise RMS at approximately 0.2 mm, 2.9% of the displacement amplitude (6.8 mm).

Although the effectiveness of the GoPro system for accurate displacement measurement has been demonstrated in this section, the undesired circumstances for vision-based systems, like camera shake and very low-contrast patterns, were fortunately avoided during the monitoring period. It is hard to ensure this satisfactory working performance for every similar monitoring exercise since the uncontrolled circumstances mentioned above often affect measurement accuracy, leading to a much higher noise level.

6.5 Displacement monitoring using a mixed system

The purpose of this section is to examine the performance of a mixed monitoring system comprising a GoPro camera and accelerometers for accurate displacement sensing under unsatisfactory conditions including considerable camera shake and poorer tracking resolution due to low-contrast patterns. This section implements the data fusion method on test data from Run 2 and 3 to evaluate the effectiveness of accuracy improvement compared to direct measurement using (only) a single camera with and without correction for camera shake.

6.5.1 Data fusion process

The GoPro measurement runs selected for data fusion involve considerable camera motion (concluded through evaluating ‘nominal’ image motions of the stationary target T00) that should be compensated in these two runs. The estimation process for extracting the mid-span displacement in Run 2 is demonstrated here. The image motions of the mid-span target T10 along the image height direction are presented in Fig. 6.10. The time history curve labelled ‘T00’ corresponds to the nominal motions of a stationary target (shifted by 0.5 pixel for clarification in the figure) located at the bridge foundation wall. The ROI T00 is expected to stay fixed during the whole recording period but actually experiences some high-frequency oscillations in both the first and last ten seconds as well as a considerable shift at 26 seconds. These effects could be attributed to the influence of camera shake and are used to correct the measurement at the ROI T10. The curve with the legend ‘T10 (corrected)’ represents the image motions of the target T10 after compensating the camera motion influence through subtracting the nominal motion of the stationary ROI T00.

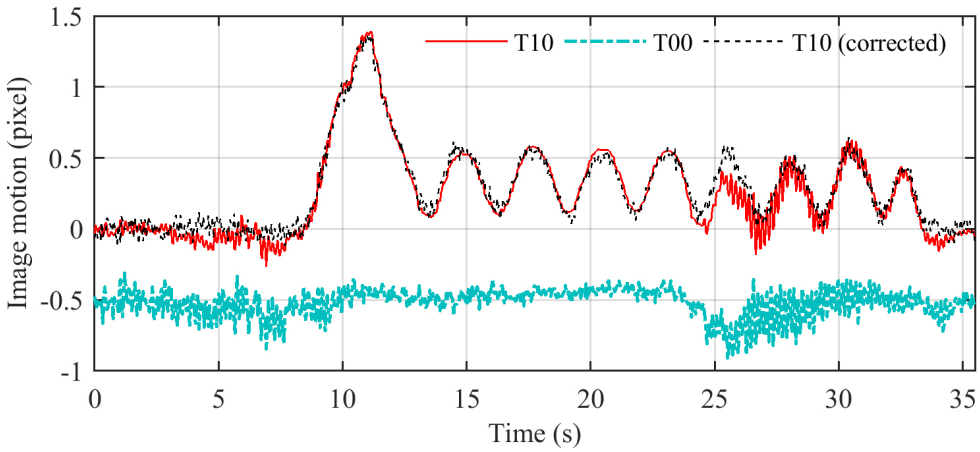


Fig. 6.10 Time histories of image motions for the target T10 along the image height direction before and after camera motion correction (curves are shifted along y axis for clarification)

Before the data fusion, the time shift between the QA accelerometer and the GoPro system is corrected by maximising the cross-correlation coefficients of the two time-history signals i.e. double-integrated displacement from the accelerometer data and the GoPro measurement after interpolation to the sample rate of accelerometer data (512 Hz). The two signals after the time synchronisation are indicated in Fig. 6.11.

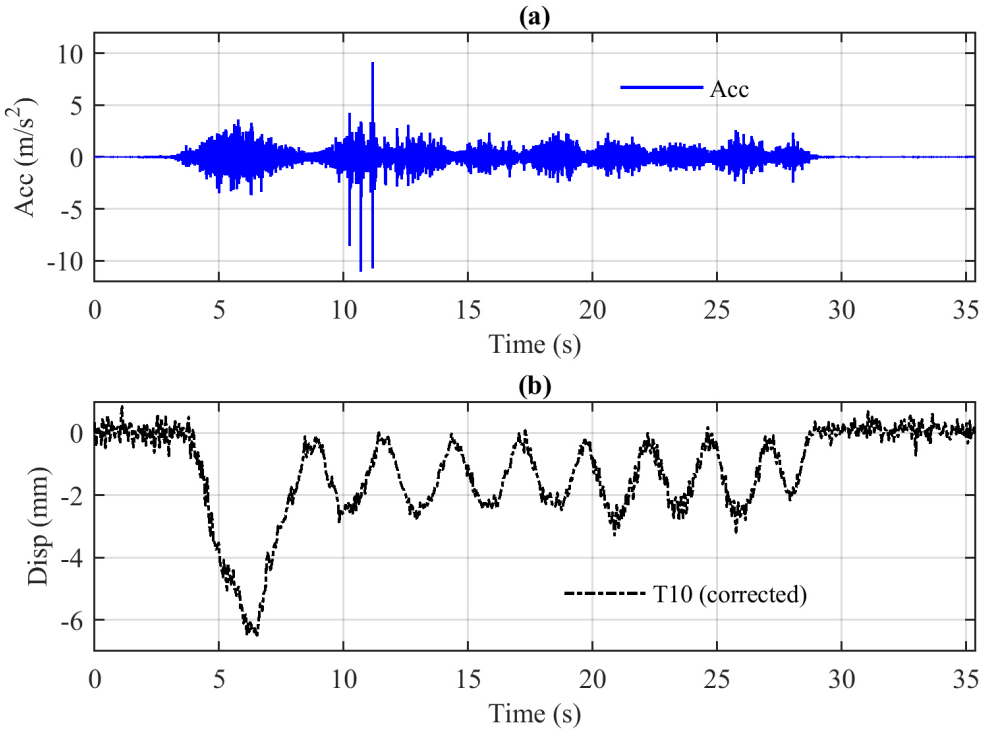


Fig. 6.11 Time histories of accelerometer and GoPro displacement data in Run 2 at mid-span in vertical direction: (a) accelerometer measurement; and (b) displacement measurement by the GoPro system

The data fusion of acceleration and displacement measurement in this study includes two main steps, the MLE for parameter tuning and Kalman filter for displacement estimation. In the MLE step, the noise variances of acceleration and displacement data are deduced through an optimisation process. For the measured data at mid-span shown in Fig. 6.11, the standard deviation of measurement noise for the accelerometer is estimated to be 0.002 m/s^2 while that for the GoPro displacement data is 0.22 mm. These results will be used in the Kalman filter estimation step to derive a refined displacement estimate.

6.5.2 Displacement estimates in Run 2

Fig. 6.12(a) indicates displacement measurement and estimates in the vertical direction for the artificial target T10 at bridge mid-span during the passage of a steam train in Run 2. The two signals with the labels of 'Raw' and 'Corrected' correspond to the displacement measurement by the GoPro system without and with camera motion compensation, respectively. The signal labelled 'Corrected + Fusion' represents the displacement estimate by fusing the acceleration data and the GoPro displacement data ('Corrected'). The maximum displacement induced by the steam locomotive is measured as 6.51 mm by the reference sensor Imetrum DMS while the maximum values in these three signals are 7.30 mm ('Raw'), 6.50 mm ('Corrected') and 6.32 mm ('Corrected + Fusion'). The cross-correlation coefficients of these three signals compared with the reference are 97.0% ('Raw'), 98.6% ('Corrected') and 99.8% ('Corrected + Fusion'). The measurement differences evaluated through subtracting the reference data (interpolated to the same sample rate) are illustrated in Fig. 6.12(b) with the RMSs at 0.53 mm ('Raw'), 0.30 mm ('Corrected') and 0.21 mm ('Corrected + Fusion'), respectively. Although the displacement data after correction ('Corrected') achieves similar value of the maximum displacement as the reference, the fusion process could effectively reduce the high-frequency noise, providing displacement data ('Corrected + Fusion') with a higher cross-correlation coefficient evaluated against the reference. The captured maximum displacement after data fusion has larger deviation (0.19 mm or 2.9%) compared with the reference but is still acceptable for the purpose of normal bridge monitoring.

Comparison results for the displacement data at T11 (the natural target at bridge mid-span) and at T20 (the artificial target at bridge one-quarter span) are presented in Table 6.3 and Table 6.4. For the target T11, the displacement estimates involving camera motion correction and data fusion process have cross-correlation coefficient of 99.7% and the measurement difference RMS of 0.24 mm evaluated against the Imetrum reference.

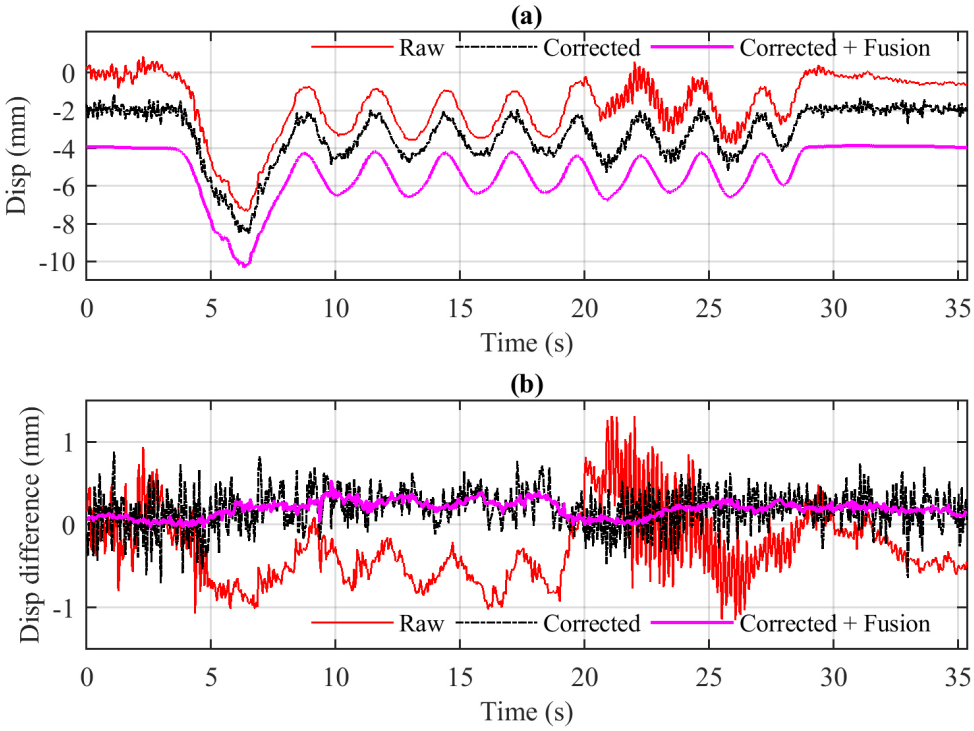


Fig. 6.12 Time histories of displacement measurement and estimates for T10 in vertical direction in Run 2: (a) displacement measurement and estimates (curves are shifted along y axis for clarification); and (b) measurement or estimation error compared with the reference Imetrum DMS

For the target T20, the displacement estimates ('Corrected + Fusion') have the cross-correlation coefficient at 99.9% and the RMS at 0.05 mm for the measurement difference evaluated by the reference measurement.

Table 6.3 Evaluation of three displacement signals for the target T11 at mid-span through comparison with the reference Imetrum DMS in Run 2

Displacement signals	Maximum displacement (mm)	Cross-correlation coefficients	RMS of difference (mm)
Reference	6.51	–	–
Raw	6.91	96.5%	0.46
Corrected	6.61	98.4%	0.35
Corrected + Fusion	6.38	99.7%	0.24

Observations from Fig. 6.12, Table 6.3 and Table 6.4 indicate that

- Camera shake could contaminate the measurement provided by a vision-based system and deserves attention in field monitoring tests. It is always preferable to implement a rigid camera mounting configuration and choose solid locations with the proper shelter for the tripod set-up.
- Camera motion correction through tracking the nominal motion of an adjacent stationary object is effective to remove the low-frequency drift, improving the

Table 6.4 Evaluation of three displacement signals for the target T20 at one-quarter span through comparison with the reference Imetrum DMS in Run 2

Displacement signals	Maximum displacement (mm)	Cross-correlation coefficients	RMS of difference (mm)
Reference	4.57	–	–
Raw	4.96	97.9%	0.29
Corrected	4.68	99.2%	0.13
Corrected + Fusion	4.57	99.9%	0.05

measurement accuracy of vision-based systems. However, the measurement resolution might be reduced when tracking the low-contrast feature target on stationary parts for correction.

- Data fusion method through fusing with the collocated accelerometer data is capable of de-noising the displacement measurement and providing better estimates about bridge displacement.

Analysis results have validated the viability of the data fusion method for improving measurement accuracy. As mention in Xu et al. (2017), another benefit of this method is to effectively widen the frequency bandwidth, which is demonstrated through Fig. 6.13. Bridge vibration signals at one-quarter span measured using the QA accelerometer in the ambient condition and in Run 2 were analysed using Welch's method to identify modal frequency information. The corresponding auto-spectral densities (ASDs) are indicated in Fig. 6.13(b) and (d). According to ambient data in Fig. 6.13(b), the first two bridge vibration modes are at 9.56 Hz and 12.50 Hz. Instead of presenting two sharp peaks as in (b), the ASD of the acceleration signal in Run 2 carries high energy in the frequency range between 8.3 Hz and 12.6 Hz and captures a sharp peak at 15.31 Hz that is not observed from the ambient result in (b). This variation is due to the varying dynamic characteristics of the coupled system of train and bridge, since the total mass of the locomotive and tender exceeds 100 t. Fig. 6.13(e) and (f) are the GoPro displacement measurement ('Corrected') and estimates ('Corrected + Fusion') in Run 2 as well as the corresponding ASDs. The GoPro displacement data with the sample rates of 24 Hz only carry an artificial mode at 5.53 Hz while displacement estimates after data fusion carry high energy near the first modal frequency (9.56 Hz) and identify a very small peak at 15.31 Hz that is less distinctive as in (d).

Analysis results indicate that the data fusion can widen the frequency bandwidth of displacement data. However, displacement data after data fusion also carry very high energy in the lower-frequency parts (e.g. $\leq$ 5 Hz) that could be misunderstood. Thus, the dynamic information of this bridge is better suited to measure using accelerometers.

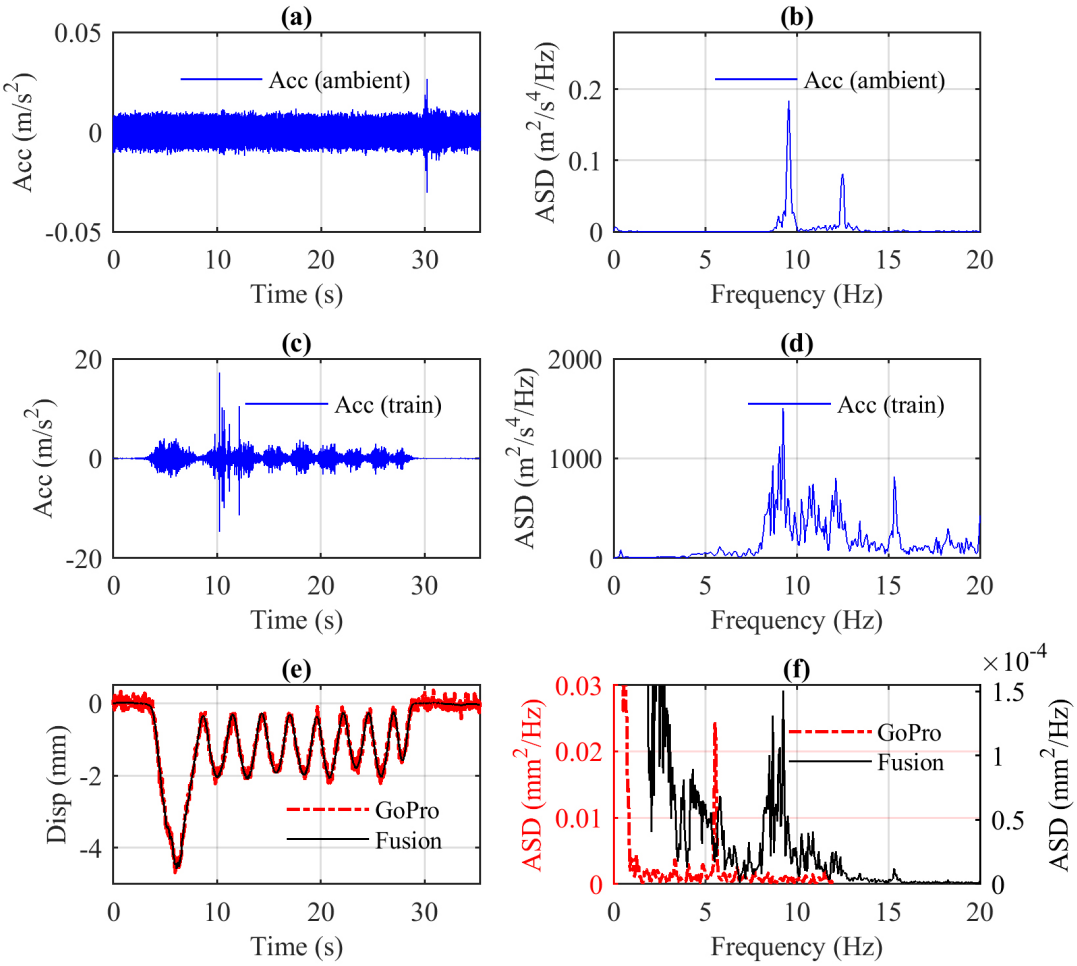


Fig. 6.13 Time series and auto-spectral densities (ASDs) of accelerometer measurement and of displacement measurement and estimates in Run 2: (a) accelerometer data in ambient condition; (b) the ASD of the data in (a); (c) accelerometer data in Run 2; (d) the ASD of the data in (c); (e) displacement data in Run 2; and (f) the ASD of the data in (e).

6.5.3 Displacement estimates in Run 3

The data in Run 2 validate the accuracy improvement for vision-based monitoring through integration with acceleration data. In fact, the direct measurement by a single GoPro system in Run 2 represents a satisfactory signal-to-noise ratio and could still capture the displacement amplitude with acceptable accuracy.

Data recorded in a more challenging test condition was chosen for further study of the working performance of the mixed system. In Run 3, apparent camera shake is observed in the recorded GoPro video and the lighting condition was poor compared with that in Run 2. Sample frames in Run 2 and 3 are indicated in Fig. 6.14(a) and (b). The ROI T00 tracked for camera motion correction has very low contrast in Run 3, which indicates a poor measurement resolution. Also, the

maximum deflection at bridge mid-span is lower than half of that in Run 2, hence requires a better accuracy for satisfactory measurement data.

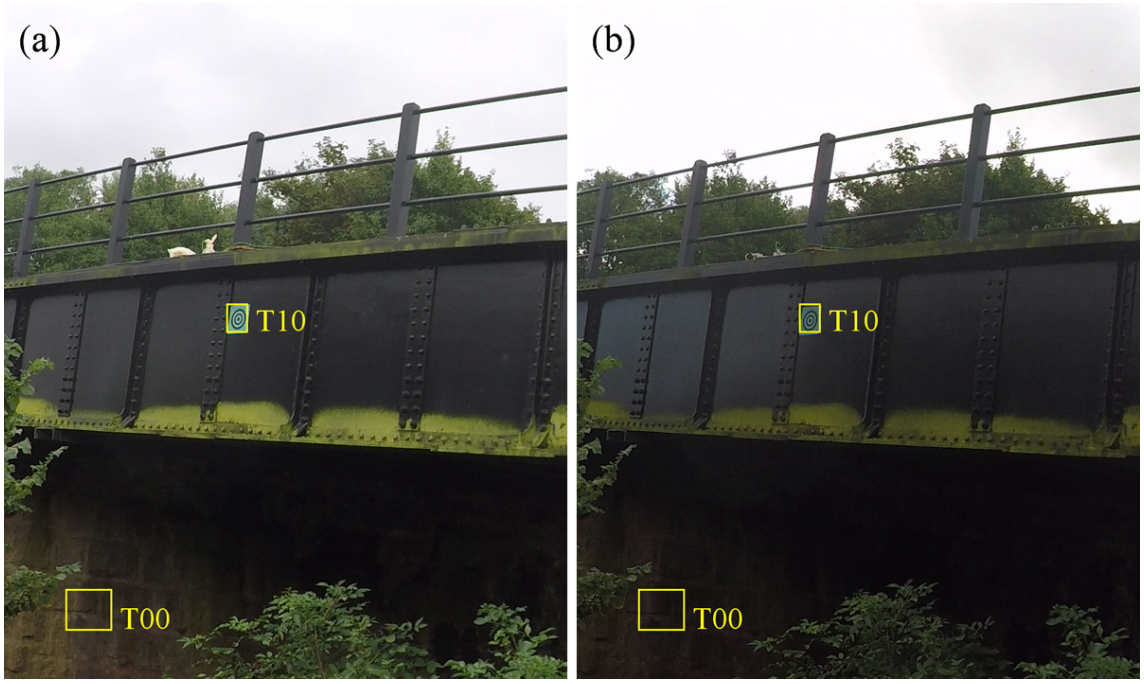


Fig. 6.14 Sample frames with marked locations of ROIs T01and T10 in Run 2 (left) and Run 3 (right)

The displacement measurement and estimates at mid-span in Run 3 are indicated in Fig. 6.15. The measurement noise is acquired by subtracting the Imetrum DMS reference data (interpolated to the same sample rate) in Fig. 6.15(b). In the raw measurement, some low-frequency drift and shaking are observed from 2 s to 5 s with the maximum deviation reaching 1 mm. Considering camera motion correction provides no improvement due to poor tracking resolution for the target T00, but instead, the RMS of the measurement difference increases from 0.34 mm ('Raw') to 0.42 mm ('Corrected'). For the 'Corrected' signal, the maximum displacement is measured as 3.93 mm, 35% higher than the reference and the cross-correlation coefficient evaluated by the reference data is 92.1%. Thus, both the displacement amplitude and the time histories measured directly by the GoPro system have poor accuracy level and are improper for bridge condition evaluation.

Through fusion of displacement ('Corrected') and acceleration data, the maximum deflection during the train passage is estimated as 2.95 mm while the reference measurement is 2.92 mm. The cross-correlation coefficient between the displacement estimate and the reference reaches 99.4% and the RMS of measured difference decreases from 0.42 mm ('Corrected') to 0.12 mm ('Corrected + Fusion').

A detailed comparison of displacement data is summarised in Table 6.5. Results indicate that the data fusion method is effective to provide accurate estimates

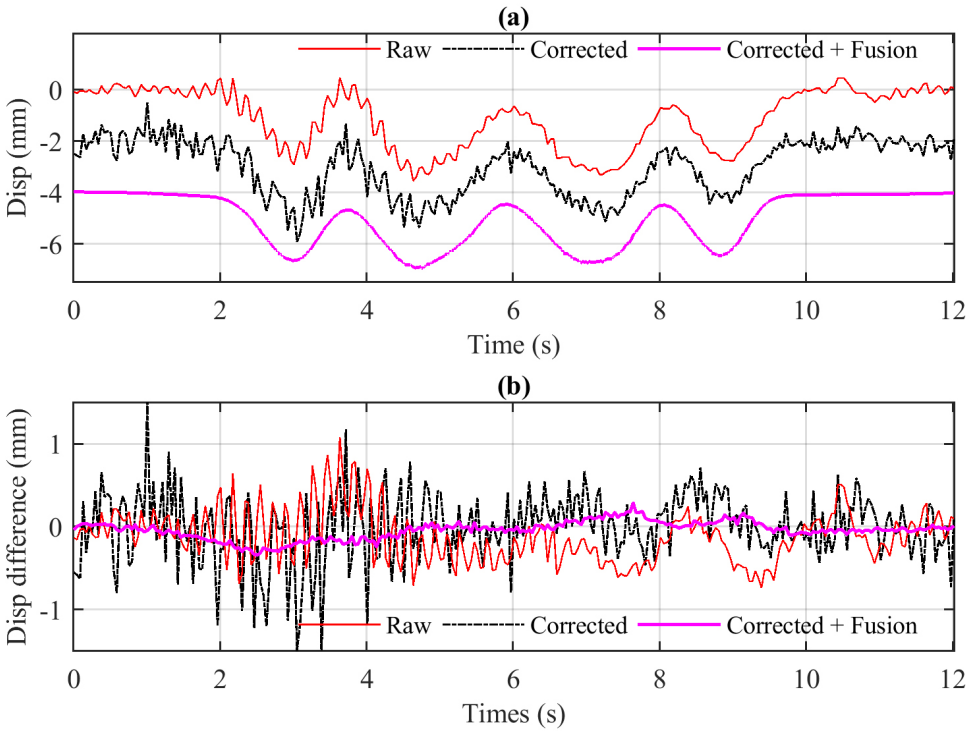


Fig. 6.15 Time histories of displacement measurement and estimates for T10 in vertical direction in Run 3: (a) displacement measurement and estimates (curves are shifted along y axis for clarification); and (b) measurement or estimation error compared with the reference Imetrum DMS

of bridge deformation even when the direct measurement from the sole camera system is unsatisfactory.

Table 6.5 Evaluation of three displacement signals for the target T10 at mid-span through comparison with the reference Imetrum DMS in Run 3

Displacement signals	Maximum displacement (mm)	Cross-correlation coefficients	RMS of difference (mm)
Reference	2.92	–	–
Raw	3.55	96.5%	0.34
Corrected	3.93	92.1%	0.42
Corrected + Fusion	2.95	99.4%	0.12

6.6 Estimation of bridge influence line

Displacement data are an important aid for bridge condition evaluation such as identifying influence lines (IL), estimating axle loads, tying into model calibration and updating, etc. This section demonstrates one application example of measured data, i.e. estimating the bridge IL.

Run 3 involving the passage of a diesel train is taken for analysis since steam trains (including a locomotive and a tender) in the other two runs have high uncertainty on weights and axle weight distributions. The diesel train (British Rail Class 115) in Run 3 includes three four-axle carriages with each carriage length of 19.50 m. The weights for the three carriages are approximately 39 tons, 30 tons and 39 tons, respectively according to manufacturer specification and are assumed to be evenly distributed to car axles for each unit. The axle locations in the front car unit are indicated in Fig. 6.16.

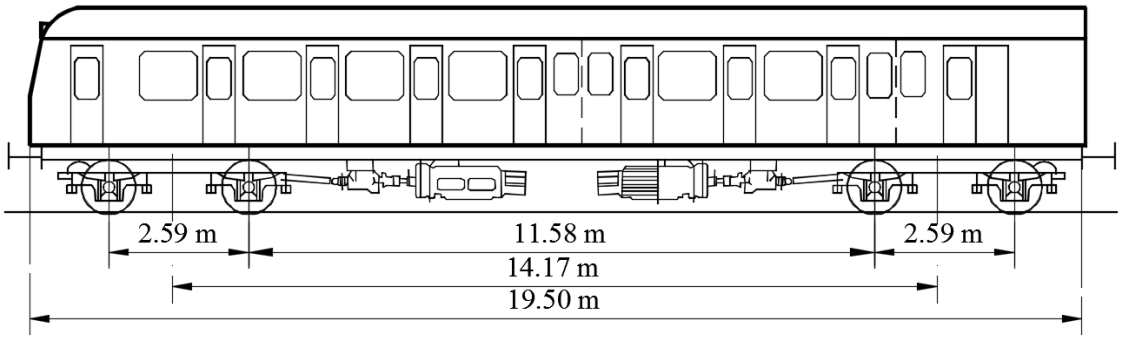


Fig. 6.16 Diagram of the first carriage in the diesel train passed in Run 3

Positions of moving axles are necessarily synchronised in time with bridge response records. From the GoPro video records, the time steps when the train front and rear passed the one-quarter span point are counted to be 2.53 s and 8.91 s, respectively. Given the total train length (60.30 m), the passing speed is estimated as 34.0 km/h, and this could be used to determine time-varying positions of each axle.

Bridge girders on two sides are of the same length (L=14.7 m) but as shown in Fig. 6.3(a), the north girder on the monitoring side is shifted back by 7.15 m along the longitudinal direction due to the bridge skew. Thus, the valid load locations X on the bridge vary from -7.15 m to 14.7 m.

Bridge displacement IL is related to time series of displacement data, train axle weights and locations expressed in a linear equation,

$$\mathbf{A}u = b \tag{6.1}$$

where b denotes time series of displacement data with the size $T \times 1$; u is a $M \times 1$ vector containing IL ordinate of the bridge that is divided into M elements (M = 50); and the matrix **A** with the dimension $T \times M$ involves the axle weight and location information and the matrix element A_{ij}corresponds to the axle load value implemented on the bridge element j at the time step i. Providing the matrices **A** and b, the IL ordinate u could be determined directly by solving $\mathbf{A}u = b$).

Displacement IL ordinate at the north mid-span is presented in Fig. 6.17. The maximal displacement at bridge mid-span under a moving unitary force (1 N) is achieved at the location X =5.61 m with the value reaching 1.33e-5 mm.

In Fig. 6.17, the considered load locations (X) in the horizontal axis of the figure are from -7.15 m to 14.7 m. The range X <0 m corresponds to the periods when the unitary force enters the south girder of this skew bridge but is not directly imposed on the north girder on the monitoring side. The bridge north mid-span starts to deform from the load position X=-2.2 m due to the cross beams linking two side girders. The IL ordinate in the range X <-2.2 m is expected to stay at zero but the IL derived from the fused displacement (GoPro fusion) actually has small drifts especially when X <-4 m which is not visible in that derived from Imetrum data. This artificial error is possibly due to small drift error remaining in GoPro displacement data even though camera motion correction is deployed.

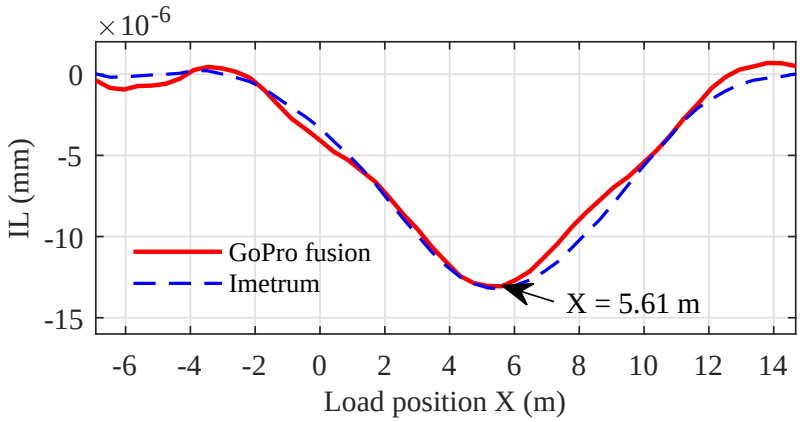


Fig. 6.17 Displacement influence line (IL) at bridge mid-span under a moving unitary force (1 N)

This section demonstrates the viability of the proposed mixed system for bridge IL estimation. One advantage of this system compared with other alternatives is that bridge responses and axle locations are derived from the same video records with no need for the time synchronisation.

6.7 Conclusion

A vision-based monitoring system based on using a single consumer-grade camera could provide accurate characterisation of bridge deformation via displacement measurement in favourable test conditions. These would include choosing salient target patterns for tracking and avoiding any camera shake. The RMS of measurement noise at the camera-to-target distance of 6.9 m is less than 0.2 mm in this example.

An effective way to correct the influence of camera shake is by tracking the nominal motion of an adjacent stationary object. This method is very effective to

remove the low-frequency drift error, but the measurement resolution is possibly reduced considerably by tracking the low-contrast feature target on stationary objects, even leading to poorer measurement accuracy after the correction. A criterion for camera stability evaluation is proposed in this study based on the tracked motions of a stationary target and the correction is performed only when necessary.

To overcome the limitation of a sole camera system, a feasible method is to fuse the vision-based displacement measurement with acceleration data for noise reduction. The data fusion method is capable of de-noising the measurement and providing better estimates of displacement. It works well even when the camera records involve apparent camera shake and low-contrast target patterns. Thus, a mixed system consisting of a camera and an accelerometer overcomes some field testing limitations of vision-based monitoring and has potential for accurate and robust displacement sensing on bridge structures.

The mixed system is demonstrated to be effective for estimating bridge influence line, indicating the application potential for bridge condition assessment.

Chapter 7

Multi-point vibration monitoring on a footbridge

Investigations in the previous few chapters focus on sensing accuracy of a vision-based system. With considerable sample rates and measurement resolutions, the system also has the potential for vibration monitoring. Compared with accelerometers, the key advantages of a vision-based system are non-contact and multi-point sensing features.

This chapter investigates the feasibility of vision-based systems for bridge vibration monitoring. The system used consists of a consumer-grade camera (GoPro) for video acquisition and a custom-developed package (reported in Chapter 4) for video processing. The system is completely non-contact overcoming the traditional dependence on artificial targets.

The monitoring structure is a cable-stayed footbridge that is easily excited under pedestrian passages. The vision-based system provides displacement measurement at multiple deck points and cables. The measurement data are implemented to extract bridge modal parameters (e.g. modal frequencies and mode shapes) and to investigate modal frequency variations under varying pedestrian loads. This application example indicates the potential of vision-based systems for non-contact vibration monitoring, providing information comparable to an array of accelerometers.

This chapter is reproduced from the paper titled "**A non-contact vision-based system for multi-point displacement monitoring in a cable-stayed footbridge**", co-authored with James Brownjohn and Dali Kong, which was published in *Structural Control and Health Monitoring* (Xu et al., 2018).

7.1 Introduction

Most of the existing applications of vision-based systems have relied on artificial targets for video processing convenience leading to necessity of direct access

to the structure as well as increased installation effort. Moreover, the focus is commonly only for single point displacement measurement e.g. at the bridge mid-span, although multi-point simultaneous sensing is supported by the camera sensors.

There have been relatively few field applications using completely non-contact vision-based systems. In most examples, an artificial target or a set of targets with salient features and some known dimensions (Martins et al., 2015; Oh et al., 2015; Stephen et al., 1993) were attached to a structure for the convenience of stable target tracking, and more importantly for providing point or line correspondences to determine the projection transformation relating the image coordinate system and the structural coordinate system. Recent non-contact field applications (Feng and Feng, 2017; Khuc and Catbas, 2017a,b; Yoon et al., 2016) have eliminated the dependency on artificial targets by using a scaling factor for camera projection transformation.

The scaling factor is the simplest method to obtain the projection transformation provided that either the camera-to-target distance or a feature dimension near the region of interest is known. The scaling factor estimated by the camera-to-target distance is sensitive to the tilt angle of the camera optical axis that is suggested to be less than 10 degrees through laboratory validation tests in short distance ($\leqslant$ 3.7 m) (Choi et al., 2016). Camera positioning is less critical for the scaling factor estimated by a known dimension (Feng et al., 2015a) but the estimated scaling factor is only reliable for displacement measurement along the same direction as the provided dimension.

Vision-based systems allow a single camera to measure structural displacements of multiple points in a structure. The feature of distributed sensing has been used in laboratory structures (Chen et al., 2015b; Oh et al., 2015; Yang et al., 2017; Yoon et al., 2016) for multi-storey displacement measurement and system identification, as well as for cable vibration monitoring (Caetano et al., 2011; Feng et al., 2017; Ji and Chang, 2008b; Kim et al., 2013) aimed at the estimation of modal frequencies or cable tensions, but applications in bridge deformation measurement are limited, with only a few examples (Busca et al., 2014; Feng and Feng, 2017).

The purpose of this study is to investigate the potential of non-contact vision-based systems for multi-point measurement in field applications. Realisation of the two features, completely non-contacting and multi-point simultaneous sensing are the focus of this study. In most applications to date, the hardware used is a professional high-resolution camera with long-focus lens, thus only a local region over the whole structure, e.g. mid-span of bridge is covered in the field of view. In this study, a low-cost consumer-grade camera with a wide angle lens is used for video acquisition with a wide area of the bridge included in the field

of view. A custom-developed package is used for the video processing which supports non-contact sensing for both deck deformation and cable vibration measurement. The developed system enables quick installation/removal, requires no access to the bridge structure and provides simultaneous multi-point displacement measurement.

To that end, section 7.2 provides descriptions of the proposed vision-based system including the hardware and video processing methods used. Section 7.3 introduces a field test in a cable-stayed footbridge under pedestrian crowd loading. Section 7.4 provides the results of the field test including the bridge deck deformation and cable vibration in time and frequency domain. The analysis results illustrate the changing bridge modal frequencies under varying pedestrian loads.

7.2 Vision-based system

The hardware of the vision-based system used in this study comprises one consumer-grade camera (i.e. GoPro Hero 4 Black) and a tripod. The recorded video files are post-processed in the custom-developed video-processing package proposed in Chapter 4.2. The measured displacement data could be interpreted for the evaluation of structural condition e.g. system identification.

Since the video processing package offers several options of sensing methodologies, section 7.2.1 and 7.2.2 present the selected methods for camera calibration and target tracking that are determined with consideration of field testing circumstances. Section 7.2.3 demonstrates the system identification methods to analyse the monitoring data.

7.2.1 Two-step camera calibration

Camera calibration is aimed at determining the transformation metric between the image natural units (pixels) and the real world units (e.g. mm), and the structural displacement could be easily derived from the change of structural coordinates given the image location of a target (output of target tracking) and a transformation metric (output of camera calibration).

Mathematically, the projection process from a 3D spatial domain to a 2D image plane loses some geometric information of the target. Thus, in a single-camera system, the calibration is realised by reducing the dimensions of target motion, i.e. assuming that the target moves within a plane in 3D space. The projection is then simplified as a 2D-to-2D transformation, enabling the recovery of the 2D structural displacement. For bridge applications, the dominant motions under traffic or pedestrian loads are in the vertical direction, making it feasible to neglect the lateral or the longitude motions in a short-time monitoring campaign.

Several methods are available to determine the transformation metric.

- Scaling factor is the simplest method based on one dimensional feature correspondence or camera-to-target distance, thus is very popular in civil applications e.g. Feng et al. (2015a); Liao et al. (2010); Stephen et al. (1993); Ye et al. (2013); Yoon et al. (2016). This method is based on an assumption that the camera principal axis is perpendicular to the structural surface plane or the two motion directions of interest, which sets constraints on camera position and orientation on site. Although the scaling factor estimated by a known dimension is less sensitive to camera positioning, the calibration should be applied separately to each target using one adjacent dimensional feature along the same direction as the movement of interest.

- Planar homography matrix is a transformation metric that links the 2D image plane with the 2D structural surface plane and is applied for the 2D motion estimation (Lee et al., 2015; Xu et al., 2016). The calibration is based on at least four sets of 2D-to-2D point correspondences (Hartley and Zisserman, 2003), i.e. structural coordinates of points in 2D structural surface plane and image coordinates of their projections in 2D image plane.

- Full projection matrix is the general form of transformation metric between the 2D image plane and the 3D structural coordinate system with no assumption, with example applications in Kim et al. (2006); Martins et al. (2015); Oh et al. (2015); Park et al. (2015). The calibration process usually comprises two steps, i) offline calibration in the laboratory to determine camera intrinsic parameters (Zhang, 2000), and ii) site calibration to estimate the camera extrinsic matrix (i.e. camera position and orientation relative to the structural coordinate system) based on at least four sets of point correspondences. The full projection matrix is the multiplication of the camera intrinsic matrix and camera extrinsic matrix.

Scaling factor is inappropriate for site applications due to the prerequisite of camera perpendicular configuration or the required dimensional features. Estimation of either planar homography matrix or full projection matrix requires some known geometric information in the structure which is commonly acquired with the assistance of some artificial targets e.g. planar targets (Chang and Ji, 2007; Kim et al., 2006) and a 3D calibration object (Martins et al., 2015). Since the offline camera calibration step in the full projection matrix method could consider lens distortion that is common in consumer-grade cameras, the full projection matrix method is used in the video processing package.

In the package, offline camera calibration is performed in the laboratory using the camera to observe a chessboard target in different views in order to obtain the camera intrinsic matrix and the lens distortion parameters. For the site calibration,

the camera extrinsic matrix is derived based on at least four sets of 2D-to-3D point correspondences. Since completely non-contact sensing is preferred, the required geometrical information is acquired from the as-built drawings e.g. the bridge span length and the pylon height. To consider the lens distortion, instead of correcting the full frame before the target tracking step, the correction occurs after the target tracking step only to the image coordinates observed from the raw frame in order to save computation efforts. Finally, the 2D structural displacement along the vertical and longitudinal directions is derived based on the corrected image coordinates and the full projection matrix.

7.2.2 Target tracking techniques

Target tracking is aimed at determining target locations in frame sequences of a video record with several techniques available:

- Correlation-based template matching is a classic and widely-used technique (Feng et al., 2015a; Liao et al., 2010; Macdonald et al., 1997; Stephen et al., 1993; Ye et al., 2013) which is realised by searching for an area in a new frame most closely resembling the reference (or template) that is pre-defined as a rectangular subset in the initial frame.

- Optical flow estimation is an established method which detects motions or flows of each pixel within the pre-determined target region based on one temporal and one spatial constraints (Sun et al., 2010). Lucas and Kanade optical flow estimation (Tomasi and Kanade, 1991) has been validated in a laboratory test of a multi-storey metal tower for system identification (Yoon et al., 2016) and applied in field monitoring of bridge stay cables during normal operation (Caetano et al., 2007).

- Feature point matching is an efficient approach that detects key-points in two images independently and then finds point correspondences based on their local appearance. Currently the applications in structural monitoring are limited, two examples being displacement monitoring test in a stadium structure using FREAK matching (Khuc and Catbas, 2017a) and in a viaduct system using SIFT matching (Khuc and Catbas, 2017b).

- Shape-based tracking is used to match special target shapes and patterns between two images, i.e. line-type target (Ji and Chang, 2008b) or custom-made targets with white and black squares (Lee et al., 2006; Lee and Shinozuka, 2006). They do not have generality for all target patterns.

Target tracking methods for the deck and cable targets are chosen separately considering their pattern features.

7.2.2.1 Tracking deck targets

Correlation-based template matching and feature point matching are the two potential approaches for tracking the deck target regions. Correlation-based template matching has been applied for structural displacement monitoring on a railway bridge (Feng et al., 2015a), a long-span bridge (Ye et al., 2013) and a high-rise building for tracking specific patterns (Chang and Ji, 2007; Macdonald et al., 1997; Stephen et al., 1993), LED lamp targets (Ye et al., 2013) and feature targets on structural surfaces (Feng et al., 2015a). One critical advantage of this method is the minimal user intervention, limited to specifying the template region in the reference frame. However, the method is sensitive to lighting changes (Brownjohn et al., 2017; Guo and Zhu, 2016) and changes of background conditions (Ehrhart and Lienhart, 2015c). Also, the method is not the ideal choice for tracking slender structural components, since a rectangular template might include background pixels that move differently from the structural elements.

Feature point matching is an alternative for the target tracking based on the key-point detection and matching. Key-points in computer vision are those that are stable, distinctive and invariant to image transformation like building corners, connection bolts or other patches with interesting shapes (Szeliski, 2011). Instead of the raw image intensities, a feature descriptor is used for matching that is a complex representation based on the shape and appearance of a small window around the key-point. Thus this technique is less sensitive to illumination change, shape change and scale variation. However, feature point matching requires the target region to have rich textures for saliency during the whole recording period. Also several threshold parameters need to be specified according to users' experience or judgement, e.g. contrast threshold for the feature detector and distance threshold in the matching criteria. These threshold values might depend on environmental conditions, e.g. the threshold for outlier removal varies with the illumination condition (Khuc and Catbas, 2017a). The existing applications are mainly focused on the short-range measurement (Ehrhart and Lienhart, 2015a,c; Khuc and Catbas, 2017a,b) while the feasibility for long-range monitoring and the stability over several hours are not validated yet.

For the studied footbridge, natural features near the bridge deck along the length direction are available for tracking, but these features are not very distinctive. The monitoring was continued over several hours, recording the different occupation states of the bridge. Thus the automatic tracking with little user adjustment under varied environmental conditions was preferred. With these considerations, correlation-based template matching is used for deck target tracking in the video processing package.

7.2.2.2 Tracking cable targets

To enable a wider field of view covering the majority of the bridge, bridge cables are projected to be slender lines in a camera image e.g. with less than four pixels along the width direction. Correlation-based template matching is inappropriate in this case since pixels within a selected template (a rectangular subset from the reference frame) might cover cable segments as well as some background (e.g. clouds and tree branches) with inconsistent motions. Optical flow estimation method faces challenges due to the limited numbers of salient feature points. The cable tracking method based on edge detection is more robust to the variations of local features and is used in the video processing package.

The cable tracking consists of two steps; edge detection and motion estimation. Edge detection is aimed at determining the cable location in a small subset window while the cable motion is estimated from the distance between two extracted edges.

In the edge detection step, a region of interest including a small cable segment is selected for tracking shown in Fig. 7.1(a). Since edge points have significant local changes in image intensity that lead to a local peak in the first derivative, image gradient is a common measure for edge detection. One of the gradient-based edge detectors, Sobel operator (Sobel, 1978) is used to detect the probable edge points (in pixel level) through calculating the image gradients among 3 by 3 neighbourhood and thresholding the magnitude of gradients. Zernike moment operator (Ghosal and Mehrotra, 1993; Ying-Dong et al., 2005) is then applied to re-located the edge precisely from the points detected by Sobel operator in Fig. 7.1(b). Zernike moments are constructed by mapping the image onto a set of complex polynomials through convolving the image intensity matrix with three pre-determined masks. Three edge parameters (i.e. step height, perpendicular distance from the mask centre and the edge direction with respect to one image axis) are estimated from Zernike moments for the probable edge points and the edge parameters are then used as criteria to remove outliers and to refine image coordinates for the remaining edge points (in subpixel level). The direction of cable segment within the selected image region is determined by line fitting among the remaining edge points in Fig. 7.1(d).

The cable motion is then estimated from the distance between two edge lines along one assumed motion direction (see Fig. 7.2). Even if the assumed motion direction deviates from the true direction, the motion estimate is proportional, so not affecting identification of cable modal frequencies.

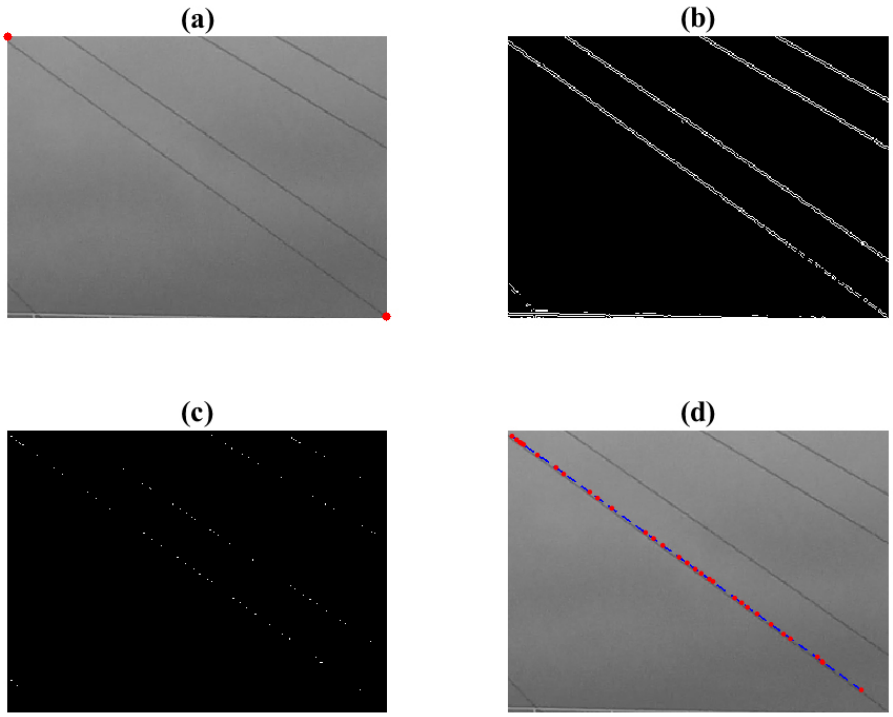


Fig. 7.1 Edge detection procedures for cable targets in the video processing package: (a) the selected cable element located at the diagonal of the cropped frame; (b) the edge detection result by Sobel operator; (c) the refined edge points by Zernike moment operator; and (c) line fitting results of one cable edge.

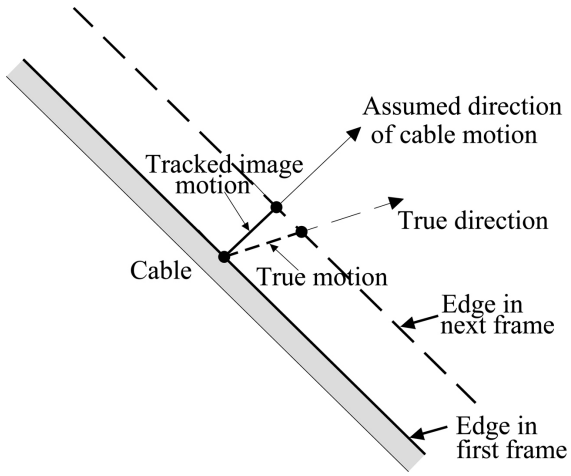


Fig. 7.2 Cable motion estimated from edge shift.

7.2.3 Monitoring data interpretation

The acquired displacement data from the video processing package could be used for data interpretation e.g. extracting the structural dynamic properties. Since the monitored structure in this study is a cable stayed footbridge loaded by crowds of passing pedestrians, the modal properties might vary with the loading.

For system identification, Welch's method (Welch, 1967) was used to estimate the power spectral densities of monitoring data by computing the average of periodograms, and the modal frequencies were estimated through peak picking. The data-driven stochastic subspace identification (SSI) method (Peeters and De Roeck, 1999) was also used to extract the modal frequencies and mode shapes through estimating a state-space model from measurement data and performing eigenvalue decomposition to the state-space model.

In time-frequency analysis, the continuous wavelet transform (CWT) was used to acquire the time-frequency distribution of measured signals including the displacement and acceleration responses. The complex Morlet wavelet was set as the mother wavelet with the relevant parameters (frequency bandwidth and central frequency) tuned according to the minimisation of the Shannon wavelet entropy (Lin and Qu, 2000; Wang et al., 2016; Yan and Miyamoto, 2006). The instantaneous frequencies were extracted from the ridges of wavelet transform modulus using the modulus maxima method (Carmona et al., 1998).

7.3 Field test on a cable-stayed footbridge

The vision-based system was applied in a monitoring test of a cable-stayed footbridge, Baker Bridge in Exeter, UK. This section described both the bridge and the configuration of vision-based system on site.

7.3.1 Bridge description

Baker Bridge is a 109 m cable-stayed footbridge crossing the A379 dual-carriageway in Exeter, UK (See Fig. 7.3). The bridge provides cyclist and pedestrian access to Sandy Park Stadium (south side of bridge), the home ground of Exeter Chiefs Rugby Club, and thus experiences heavy pedestrian traffic on match days. The bridge comprises a single A-shaped tower that supports the continuous steel deck over a simple support at the pylon cross-beam as well as via seven pairs of stay cables.

In a previous ambient modal test (Brownjohn et al., 2016a), four modal frequencies below 2.5 Hz were observed in the vertical direction, i.e. 0.94 Hz, 1.62 Hz, 2.0 Hz and 2.24 Hz. Thus, the bridge has noticeable vibration response due to pedestrian traffic.

7.3.2 Description of a monitoring test

A four-hour monitoring test was performed from 13:37 to 17:42 on the afternoon of the Exeter Chiefs vs. Newcastle Falcons fixture on 25th Feb 2017, with a crowd

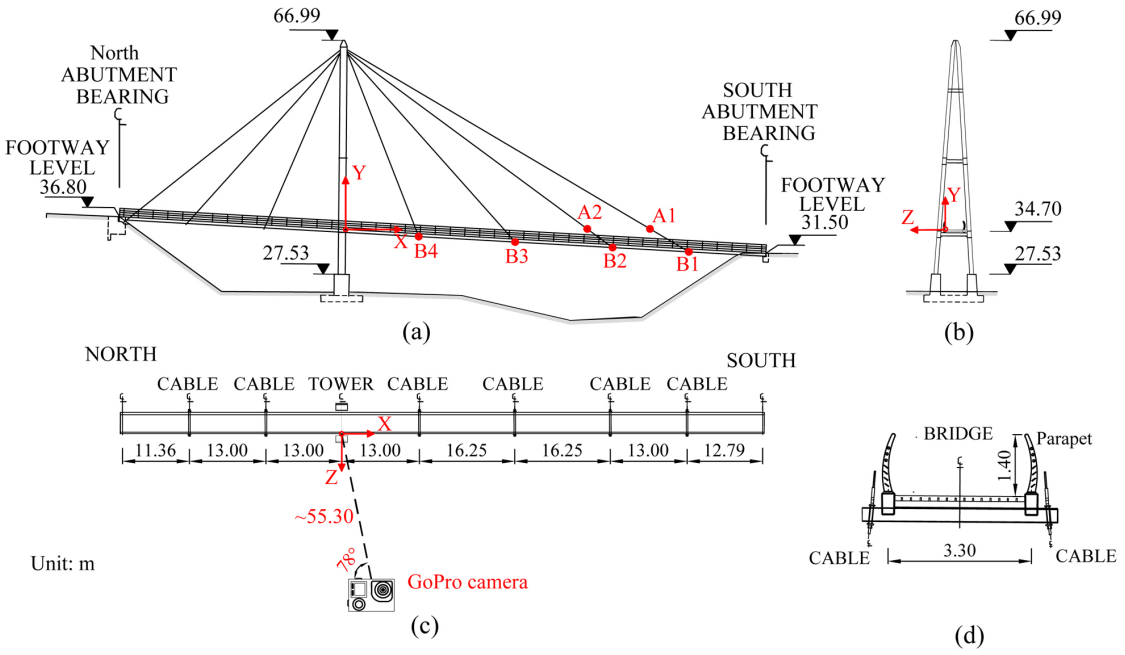


Fig. 7.3 Baker Bridge information and the sensor locations: (a) the west elevation of the bridge with the marked locations of six wireless accelerometer sensors APDM OpalTM (B1∼4, A1∼2); (b) the tower elevation of the bridge; (c) the deck plan of the bridge with the marked location of the GoPro camera in the west side; and (d) a typical deck section of the bridge.

of 10,469 spectators (Exeter Chief News, 2017), many of them arriving by train (Digby and Sowton Station) and crossing the bridge to reach the stadium. Match kick off was at 15:00 with two halves of 40 minutes each plus stoppages and a 10 minute interval. A GoPro Hero 4 Black camera, Fig. 7.3(c), was mounted on the top of a tripod at the central reservation of the A379 carriageway below and approx. 55.30 m southwest of the bridge tower. The west side of bridge was included in the field of view with one sample frame indicated in Fig. 7.4(a). The nominal frame rate was set as 30 Hz while the actual one was 29.97 Hz. Narrow field of view setting was selected with the corresponding focal length equivalent to 30-34 mm. The image dimensions was 1920 pixels × 1080 pixels. The weather was overcast with little illumination change during recording.

The video processing procedures consist of three main steps similar to those in the laboratory validation test. During the camera calibration, camera intrinsic parameters were determined ahead of the test. A sample corrected frame after removing the influence of lens distortion is shown in Fig. 7.4(b). Camera extrinsic matrix was determined on site based on several pairs of 2D-to-3D point correspondences. The structural coordinate system was specified with the origin at the deck height of the tower section, the Y axis along the vertical direction and the Z axis along the transverse direction (See Fig. 7.3(a) and (b)). The control points (CP) are marked in Fig. 7.4(b) with known structural coordinates from the as-built drawings provided by the Devon County Council: CP1∼4 along the edge of the

bridge tower and CP5∼11 near the outer-section of the crossbeams to which the cables are secured.

For target tracking, four targets (D1∼4) along the deck longitude direction and two targets (C1∼2) at the cable edges were chosen for tracking, all at the southwest side of the bridge in Fig. 7.4(a). The pixel dimensions of the selected targets in video frames are indicated in Table 2. Due to the limited availability of stable features in the bridge deck, the height of these deck targets is approximately 20 pixels, smaller than the suggested value (40 pixels) in the previous study (Brownjohn et al., 2017).

For deck targets (D1∼4), the structural displacement along the longitude and vertical direction was estimated based on camera calibration and target tracking results. For the cable targets (C1∼2), the cable motion estimated in the target tracking step was directly outputted.

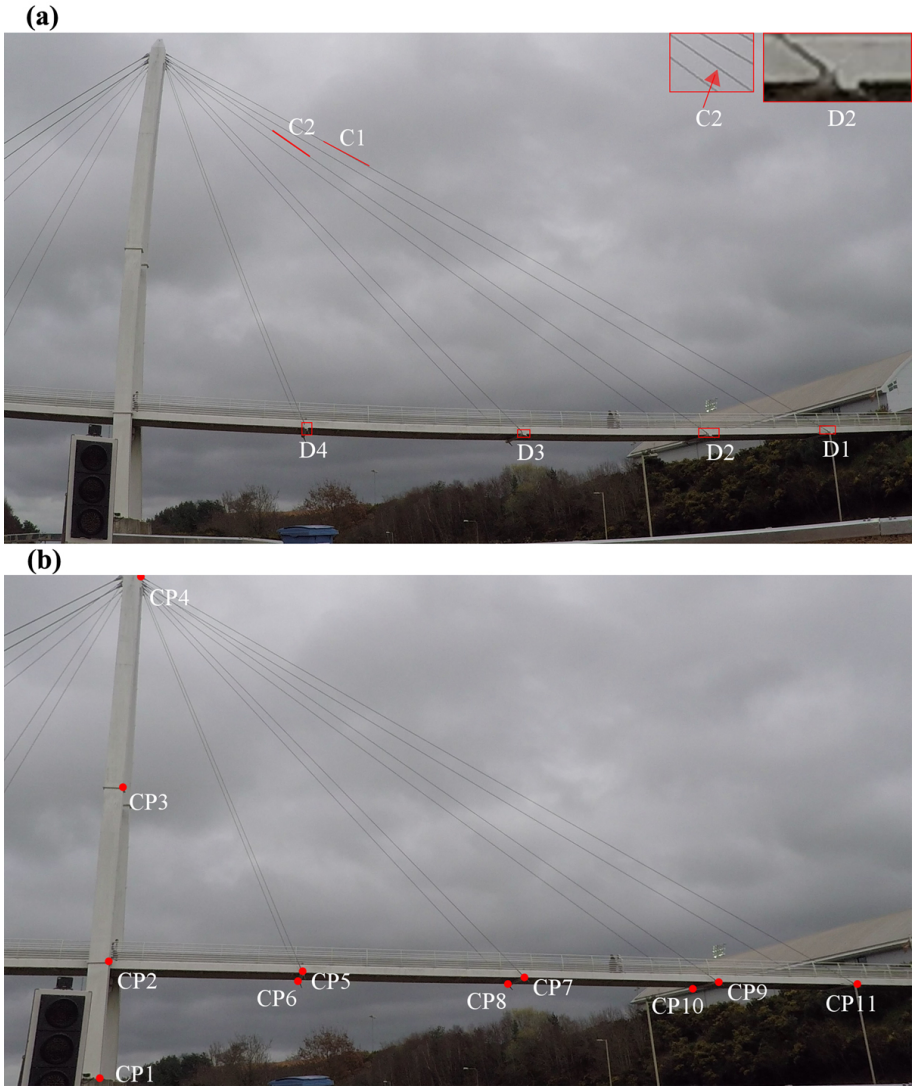


Fig. 7.4 Sample frames from the video records: (a) one raw frame (including apparent lens distortion) with markers on deck and cable targets for video tracking; and (b) one corrected frame (removing lens distortion influence) with markers on control points (CP1∼11) used for the calibration of camera extrinsic matrix.

As well as the vision-based system, six tri-axial wireless accelerometers (APDM OpalTM) were installed on the bridge: four (B1∼4) on the deck parapet and two (A1∼2) on the cables with locations marked in Fig. 7.3(a). The purpose of the Opal sensors was to corroborate the identification of modal parameters of bridge deck and cables obtained using the vision-based system. The Opal sensors B1∼4 corresponded to the target regions D1∼4 in the vision-based system while the sensors A1∼2 were collocated to the same cable sequences as the target regions C1∼2. The sample rate was set to 128 Hz.

7.4 Measurement and analysis results

In this section, the measurement results obtained by the vision-based system are illustrated in time and frequency domains. The time interval for analysis from 16:39 to 17:14 (35 min) thus included periods when large crowds of spectators crossed the bridge on the way home after the match. The measured data from the vision-based system was analysed to investigate the dynamic properties of the bridge including the changing modal frequencies under varying pedestrian loads. Section 7.4.1 and 7.4.2 demonstrate the measurement and analysis results of bridge deck displacement and cable vibration, respectively.

7.4.1 Measurement and analysis of deck displacement

The vertical displacements of the four deck targets along the bridge span are described in this section. The measured data are presented in time domain (section 7.4.1.1) and frequency domain (section 7.4.1.2), respectively.

7.4.1.1 Time history measurement of vertical displacement

Four deck targets (D1 4) were tracked with the time histories of vertical displacement shown in Fig. 7.5 and four extracted frames from the video files in Fig. 7.6. During the recording, the bridge changed from almost empty (Fig. 7.6(a)) to almost full (Fig. 7.6(b)), then reverting to a trickle of pedestrians (Fig. 7.6(d)).

In Fig. 7.5, an obvious downward trend of the bridge deck is observed from 800 s to 1250 s in the measured data at D1 and D2 with the maximum deformation value reaching 72.58 mm and 64.10 mm, respectively. A quick deformation recovery is seen at approximately 1300 s from the measurement at D1 D3 that should correspond to a sudden reduction in bridge loading. The captured frame at 1315 s (Fig. 7.6(c)) shows a clear gap (approx. 16.5 m) between two groups of pedestrians, which accords with observations from the measured data.

Frequent local sharp peaks are observed only in the vertical displacement of D1 in Fig. 7.5. Fig. 7.7(a) zooms into 60 s of signal, and inspection of the video

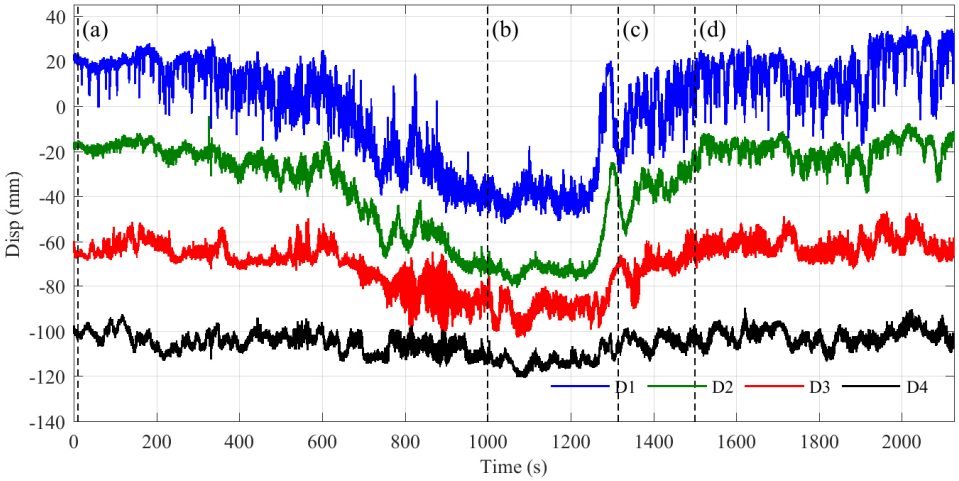


Fig. 7.5 Time histories of vertical displacement at four deck targets D1∼D3 by the vision-based system.

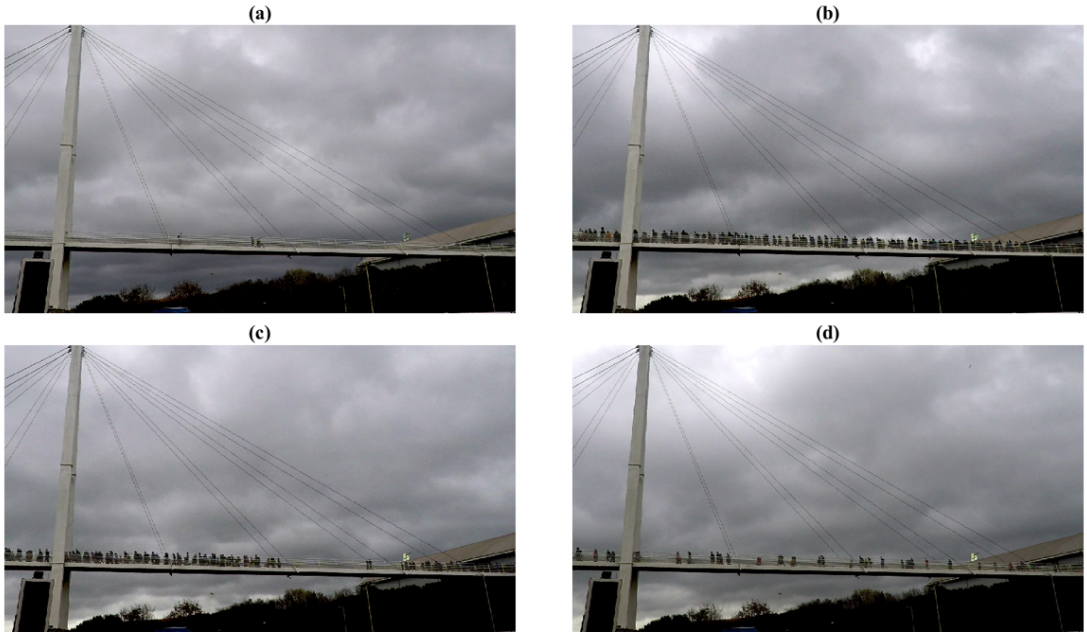


Fig. 7.6 Four captured frames for the recorded video files corresponding to the time step: (a) at 10 s; (b) at 1000 s; (c) at 1315 s and (d) at 1500 s.

recording shows that the sharp peak of 14.89 mm in the time interval from 34 s to 41 s was induced by the passing of two small groups of pedestrians from the opposite directions shown in Fig. 7.7(b) and (c).

7.4.1.2 Frequency components of vertical displacement

The power spectral densities (PSD) of vertical displacement and acceleration measurement were estimated using Welch's method with the window length of one minute duration and a 50% overlap. Fig. 7.8 illustrates the estimation results for the signals recorded during three time intervals i.e. ([0, 400] s, [800, 1200] s and [1600, 2000] s). The first time interval ([0, 400] s) was at the end of the match

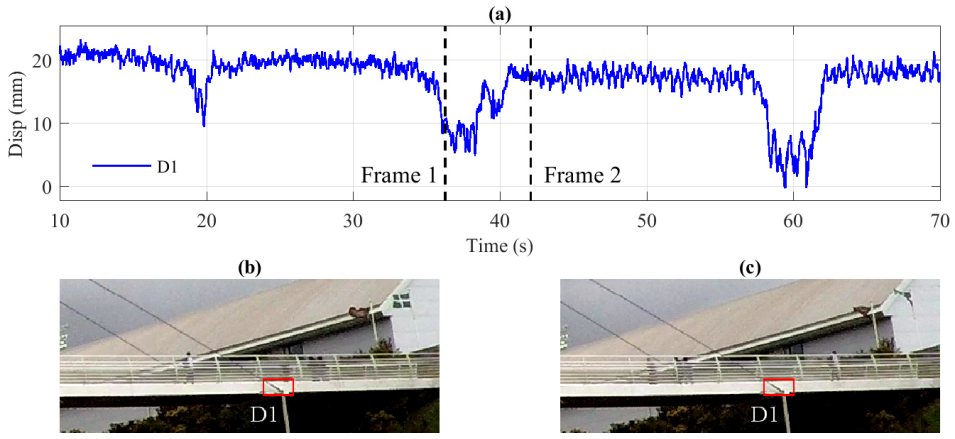


Fig. 7.7 A 60-s time history signal of deck displacement in vertical direction at the target D1 with passing pedestrians: (a) time history of vertical displacement at D1; (b) Frame 1 at 36 s; and (c) Frame 2 at 42 s.

(during stoppage time) and thus few pedestrians crossed the bridge; the second duration ([800, 1200] s) was after the Rugby match and the bridge was almost fully occupied by pedestrians; and the third time interval was after most spectators had left and still a few pedestrians were crossing the bridge. Acceleration data of the deck at B3 was not available due to a faulty battery.

In Fig. 7.8(a), four apparent modal frequencies are identified using peak-picking with the values of 0.92 Hz, 1.61 Hz, 2.00 Hz and 2.23 Hz, which match well with the results from acceleration measurement in (b). The displacement measurement at the deck point D1 contains significant quasi-static response due to the local deformations resulting from passing pedestrians (See Fig. 7.7), preventing identification of the first modal frequency at 0.92 Hz.

In the second time interval shown in (c), only the second modal frequency is clearly indicated, with the value decreased to 1.48 Hz (from 1.61 Hz). The signal power near this frequency value is increased sharply compared with the data in the other two periods. The shift of the second modal frequency is also observed in (d) from acceleration data.

In the third time interval shown in (e), the second mode still contains the highest power with frequency value shifted back to 1.59 Hz. The first and third modal frequencies are identified with the same values (0.92 Hz and 2.00 Hz) as in the first time interval. The observations match well with the analysis results of the acceleration data shown in (f).

Through the analysis, it indicates that:

- The measured data by vision-based system captures the modal frequencies of the bridge deck accurately through the comparison with the acceleration data;

- The second mode of the bridge deck is very sensitive to the occupation status of the bridge and the frequency value reduced from 1.61 Hz to 1.48 Hz with full pedestrian occupancy, corresponding to a reduction of 8%.

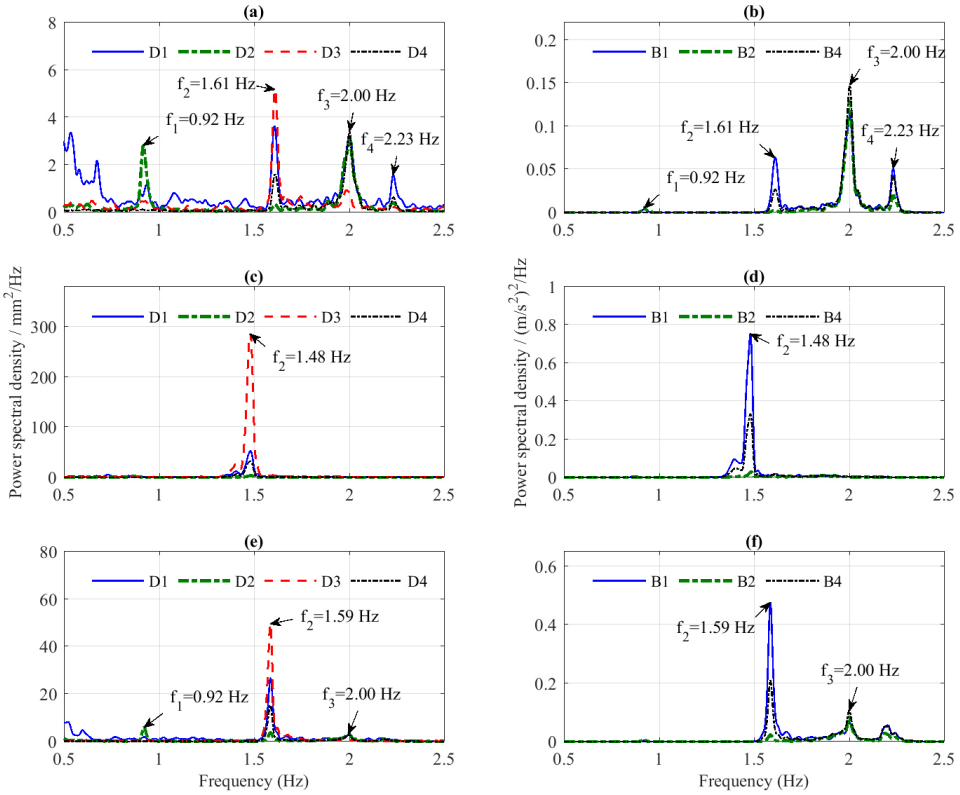


Fig. 7.8 Power spectral densities (PSD) of time history signals with the three rows representing the three time intervals, the left column representing the PSD of displacement measured by vision-based system and the right column representing the PSD of acceleration measured by Opals: (a) PSD of vertical displacement at the time range of [0, 400] s; (b) PSD of vertical acceleration at the time range of time range of [0, 400] s; (c) PSD of vertical displacement at the time range of [800, 1200] s; (d) PSD of vertical acceleration at the time range of [800, 1200] s; (e) PSD of vertical displacement at the time range of [1600, 2000] s; and (c) PSD of vertical acceleration at the time range [1600, 2000] s.

The SSI method (Peeters and De Roeck, 1999) was used to identify the modal frequencies and mode shapes from the collected data in the third time interval (i.e. [1600, 2000] s) and the analysis results were compared with those observed from a previous ambient modal test using APDM opal sensors (Brownjohn et al., 2016a). Fig. 7.9 compares results for the two bending modes in vertical direction:

- The second modal frequency estimated by displacement data is 1.58 Hz, lower than the value (1.62 Hz) reported in (Brownjohn et al., 2016a). This is due to more frequent crossing pedestrians on the test day.

- The third modal frequency (2.00 Hz) estimated by displacement data matches the value in (Brownjohn et al., 2016a).
- For these two bending modes, the mode shape ordinates (red circular dots) at the points D1 4 predicted by the vision-based measurement match well with the mode shapes (blue curves) previously estimated in (Brownjohn et al., 2016a).

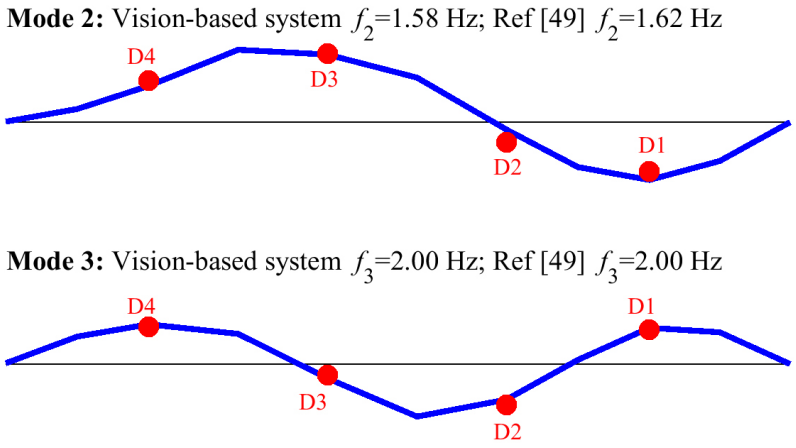


Fig. 7.9 Mode shapes and frequency estimates of the bridge longer span: blue curves represent the mode shapes extracted from the previous ambient modal test (Brownjohn et al., 2016a) corresponding to the longer span closest to the stadium; and red dots represent the mode shapes extracted from displacement data measured by vision-based system.

The frequency responses of measured signals in Fig. 7.8 indicate dependency on time, whose study requires time-frequency analysis rather than methods based on the Fourier transform (e.g. Welch's method) that are designed for the analysis of stationary signals. Continuous wavelet transform (CWT) analysis was therefore used to acquire the time-frequency distribution of displacement and acceleration measurement.

Fig. 7.10(a) indicates the CWT results for displacement measurement at the deck target D1 with the frequency range from 1.3 Hz to 1.8 Hz which covers the variations of the second modal frequency of the bridge deck (varying from 1.48 Hz to 1.61 Hz in Fig. 7.8). During the analysis, the two parameters (frequency bandwidth f_b and central frequency f_c) in the complex Morlet wavelet were tuned according to the minimisation of Shannon wavelet entropy, reaching optimal parameters at f_b = 4.5 Hz and f_c = 29 Hz. To consider the edge effect, the influenced region was estimated to be 40 s duration according to Yan and Miyamoto (2006) and the padding scheme of reflecting the signal at two ends was used to mitigate the edge effect. A threshold (e.g. -0.5 in Fig. 7.10(a)) was set for the wavelet transform modulus value during the plotting for a clear visualisation. The instanta-

neous frequencies were estimated by the modulus maxima at each time step and are shown as the sparse dots in the figure. The results in Fig. 7.10(a) indicate an obvious variation of the second modal frequency during the recording period.

- In the first 500 s, the modal frequency has small deviation with the value over 1.60 Hz.
- During the time interval from 600 s to 1100 s, a sharp decrease of the modal frequency value is observed with the lowest shifting to approx. 1.37 Hz, a reduction of 15%.
- The data after 1500 s reflects a recovery of frequency value to approx. 1.58 Hz.
- These observations match well with the analysis results for acceleration data (B1) shown in Fig. 7.10(b).

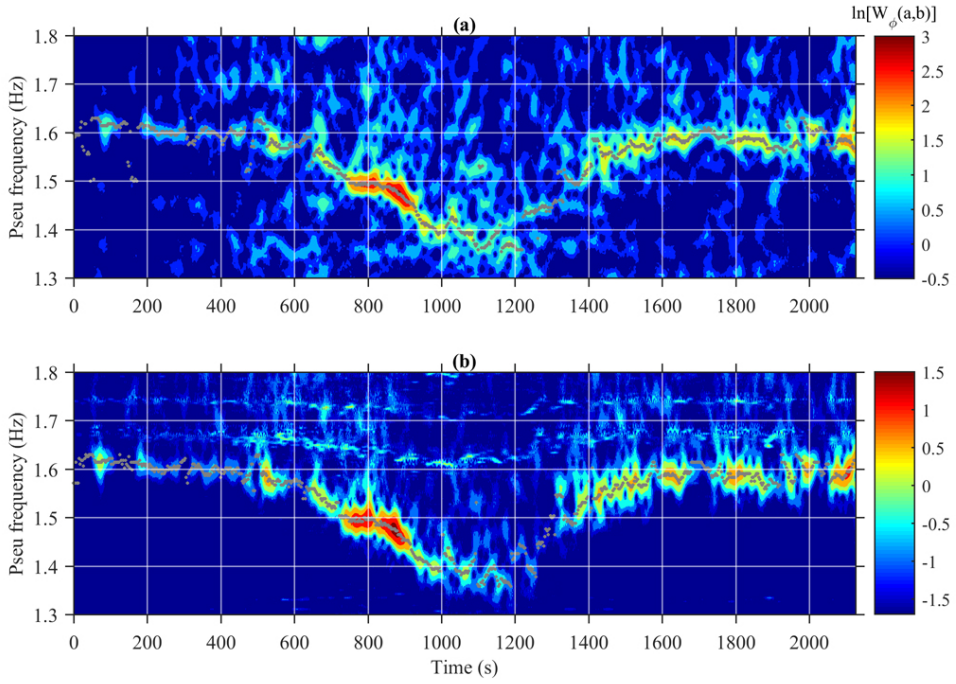


Fig. 7.10 Contour plot of CWT analysis results of displacement measurement (D1) and acceleration measurement (B1): (a) wavelet transform modulus for the measured displacement by vision-based system at the frequency range of [1.3, 1.8] Hz with the estimated instantaneous frequencies marked as sparse dots; and (b) wavelet transform modulus for the measured acceleration at the frequency range of [1.3, 1.8] Hz with the estimated instantaneous frequencies marked as sparse dots.

7.4.2 Measurement and analysis of cable vibration

This section presents the measurement results of cable vibration using the vision-based system. The measured data were directly used to estimate cable modal

frequencies by peak-picking from power spectral densities as well as using SSI. To evaluate the variations of cable modal frequency with changing pedestrian loads, the CWT analysis was performed on the measured data to identify the time-frequency distribution of cable vibration that was compared with the observations from acceleration measurement.

Two cable targets C1 and C2 shown in Figure 11(a) were tracked, with the time histories of cable motion shown in Fig. 7.11(a) and (c). The cable motion here corresponds to the motion of cable projection in the image in pixel units. The power spectral densities of the cable motions during the three time intervals (i.e. ([0, 400] s, [800, 1200] s and [1600, 2000] s) are indicated in Fig. 7.11(b) and (d) for the cable targets C1 and C2, respectively.

- The modal frequency of the cable C1 is approximately 1.66 Hz for the first and third time intervals and slightly increased for the second time window.
- For the cable C2, the modal frequency could be identified as approximately 2.10 Hz during the first and third time intervals while the analysis result for the data from the second time interval indicates no obvious peak frequency, but rather a frequency range with higher energy near 2.2 Hz.

The SSI method was used to identify the modal frequencies from the collected displacement and acceleration data during the third time interval (i.e. [1600, 2000] s). Cable target C1 (in Figure 11(a)) and accelerometer A1 (in Figure 10) correspond to the same bridge cable (the longest one, in the southwest side) while the sensor locations were different. C1 was at approximately 1/4 span point close to the bridge tower and A1 was in the lower height close to the cable end where it is attached to the bridge deck. Similarly, C2 and A2 correspond similarly.

The displacement signal at C1 indicates two close modal frequencies at 1.59 Hz and 1.66 Hz while the acceleration signal at A1 captures the first modal frequency at 1.63 Hz as well as the higher modal frequencies at 4.96 Hz, 6.62 Hz and 8.27 Hz. Through comparison, the fundamental frequency of the longest cable (C1) was at approximately 1.66 Hz. The mode at 1.59 Hz identified from the displacement signal might correspond to the second bending mode of the bridge deck. The first modal frequency estimated from the acceleration data is different from the estimated fundamental frequency, which might be due to mixing of frequency responses between the cable fundamental mode and the second bending mode of bridge deck.

The displacement signal at C2 indicates modal frequencies at 2.10 Hz and 4.16 Hz while the acceleration data at A2 capture modal frequencies at 2.12 Hz, 6.30 Hz, 8.36 Hz and 10.48 Hz. Therefore, the fundamental frequency of the second longest cable (C2) was at approximately 2.10 Hz.

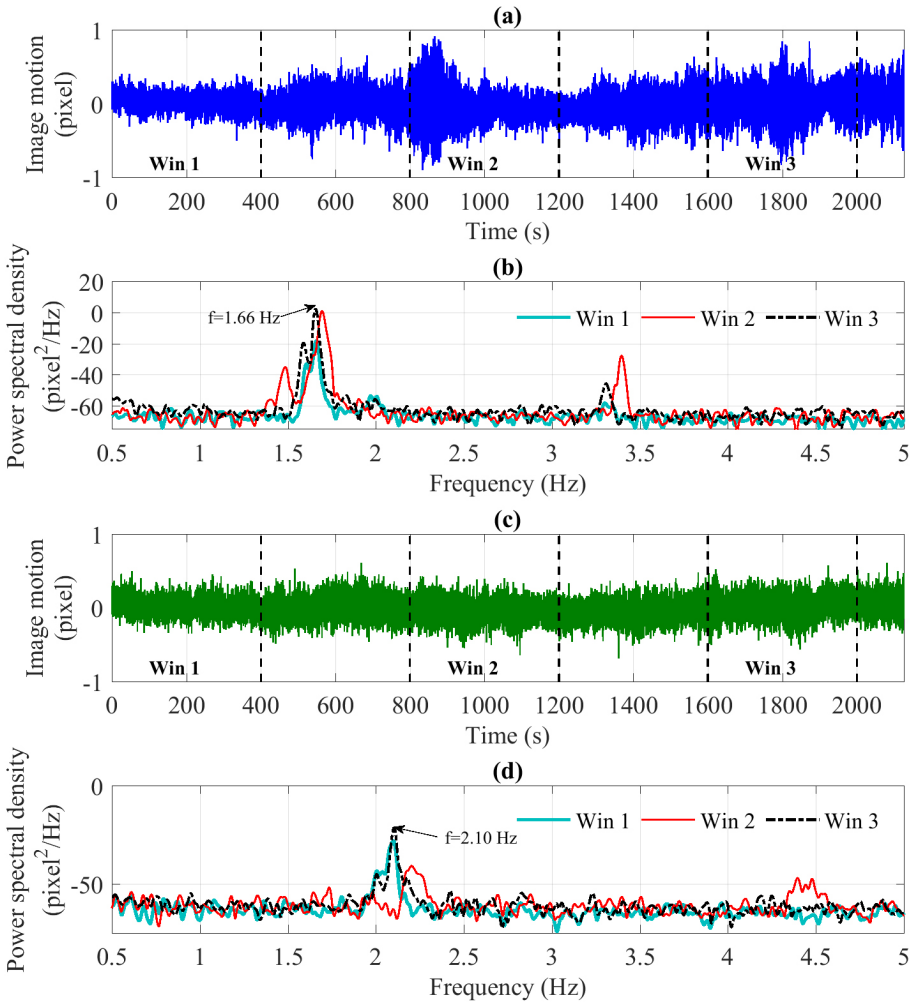


Fig. 7.11 Measurement of cable motions by the vision-based system: (a) image motion at C1; (b) power spectral densities of image motion at C1 in three marked time windows shown in (a) ([0, 400] s, [800, 1200] s and [1600, 2000] s); (c) image motion at C2; and (d) power spectral densities of image motion at C2 in three marked time windows shown in (c) ([0, 400] s, [800, 1200] s and [1600, 2000] s).

The analysis indicates that the vision-based system works better to capture the lower modal frequencies of cables while the accelerometers provide reliable estimations of higher frequency modes.

CWT analysis was performed to acquire the time-frequency distribution of cable vibrations from the vision-based system and accelerometers. The analysis results are shown in Fig. 7.12 and Fig. 7.13 for the cables C1 and C2, respectively. In terms of accelerometer measurement, instead of plotting directly the results near the fundamental frequency of the cable, higher frequency ranges i.e. near the fifth modal frequency for the measurement at A1 and near the third modal frequency for the measurement at A2 are illustrated in Fig. 7.12(b) and Fig. 7.13(b), with the corresponding values near the fundamental frequency marked in the right y-axis.

Fig. 7.12(a) indicates the CWT analysis results for the measurement at the cable C1 by vision-based system. The instantaneous frequencies estimated by the modulus maxima (shown in the figure as sparse dots) initialled at approximately 1.66 Hz, rose to over 1.70 Hz during the time interval from 900 s to 1160 s and then recovered to 1.66 Hz after 1500 s. Since the time interval from 900 s to 1160 s corresponds to the period where the deck points D1 and D2 experienced a large deformation (See Fig. 7.5), the observations indicate that heavy pedestrian loads on the bridge lead to a rise in cable modal frequency by 2.4%, probably by increasing the cable tension. Compared with the analysis result of acceleration measurement shown in Fig. 7.12(b), the time-frequency distribution acquired by vision-based measurement captures the general trend of frequency shift under pedestrian loads over the whole 35 min. However, some details of frequency variation within a short-time range are only identified by acceleration measurement, e.g. a sharp decrease and recovery of cable modal frequency at approx. 1280 s.

As well as the cable C1 modal frequency at approx. 1.66 Hz, another less obvious mode is indicated in Fig. 7.12(a) with frequency value lower than the cable fundamental frequency. This mode is salient in the lighter loading condition e.g. (i) in the time interval from 200 s to 900 s with the modal frequency decreasing from 1.62 Hz to 1.46 Hz; and (ii) the time interval from 1400 s to 2100 s with the modal frequency increasing from 1.53 Hz to 1.60 Hz. The observed mode shows a similar trend as the variation of the second modal frequency of the bridge deck. Therefore, this mode might be due to forced vibration of the cable by motion of the bridge deck.

The CWT analysis results for the measurement by vision-based system at the cable C2 are indicated in Fig. 7.13(a) with the estimated instantaneous frequencies shown as sparse dots. The cable modal frequency started off at approx. 2.08 Hz when the bridge was occupied by only a few pedestrians. An obvious rise of modal frequency (exceeding 2.2 Hz) is observed during the time range from 900 s to 1260 s when the bridge was under heavy pedestrian loads. Compared with the quiet period, the maximum shift of cable modal frequency reaches 9.1% with the frequency value reaching 2.27 Hz. In the period after 1500 s, the modal frequency of cable vibration was recovered to approx. 2.1 Hz. These observations match well with those from Fig. 7.13(b) corresponding to the acceleration measurement (A2) of the same cable. However, the analysis results of acceleration data illustrate better resolution of the variations of cable modal frequency with time, especially during heavy load periods from 900 s to 1260 s.

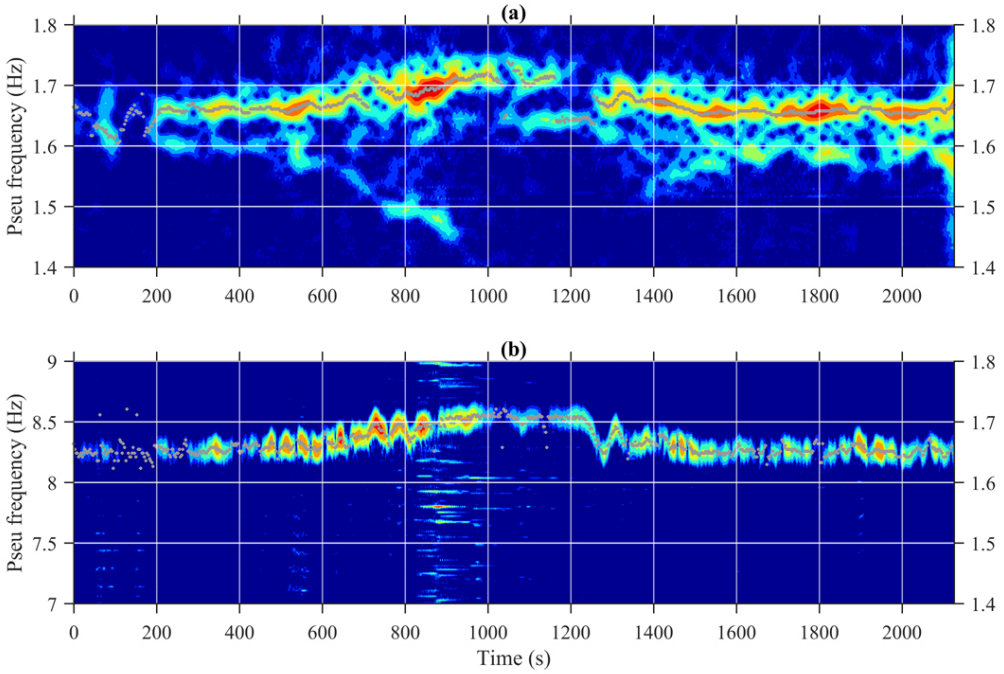


Fig. 7.12 Contour plot of CWT analysis results of cable vibration for the longest cable in the southwest side of the bridge: (a) wavelet transform modulus for the cable motion (C1) measured by vision-based system at the frequency range of [1.4, 1.8] Hz with the estimated instantaneous frequencies marked as sparse dots; and (b) wavelet transform modulus for the cable vibration measured by the accelerometer (A1) at the frequency range of [7, 9] Hz with the estimated instantaneous frequencies marked as sparse dots.

7.5 Conclusions

A non-contact single-camera vision-based system used for non-contact measurement of bridge displacement provides results comparable to those obtained using an array of wireless accelerometers and offered additional information about quasi-static response to varying pedestrian loads.

Vision-based measurement provides accurate estimation of modal parameters (e.g. bridge modal frequencies and mode shapes). The data could also be implemented to study modal frequency variations under varying pedestrian loads.

The multi-point deformation data obtained using the vision system proved to be effective for tracking cable dynamic properties at the same time as bridge deformation, allowing for the effect of varying load on cable tensions to be observed. This provides a powerful diagnostic capability for larger cable-supported structures.

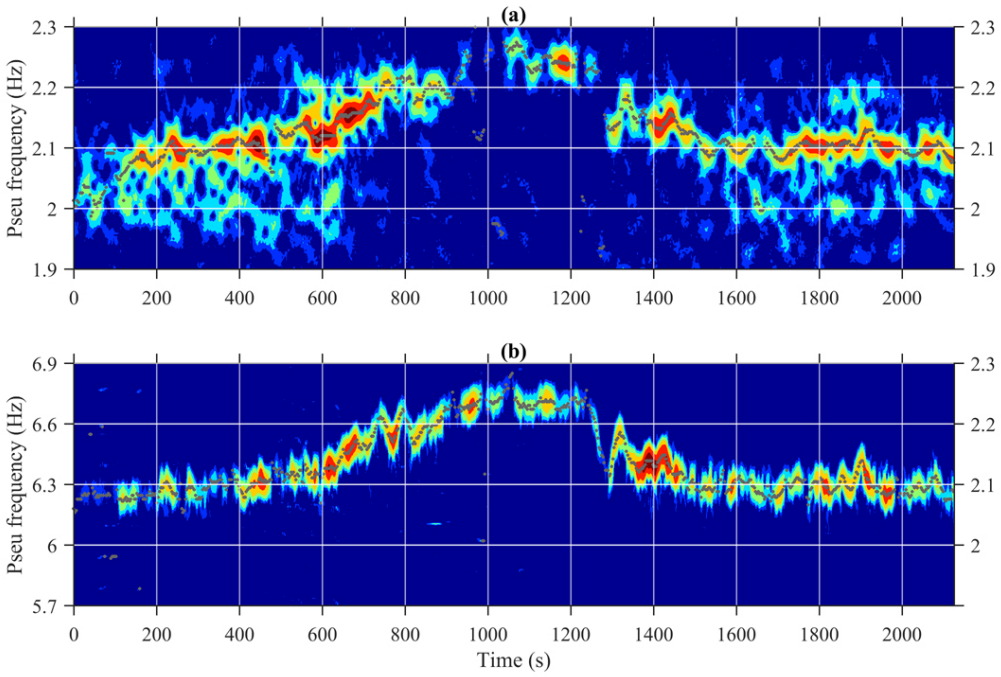


Fig. 7.13 Contour plot of CWT analysis results of cable vibration for the second longest cable in the southwest side of the bridge: (a) wavelet transform modulus for the cable motion (C2) measured by vision-based system at the frequency range of [1.9, 2.3] Hz with the estimated instantaneous frequencies marked as sparse dots; and (b) wavelet transform modulus for the cable vibration measured by the accelerometer (A2) at the frequency range of [7, 9] Hz with the estimated instantaneous frequencies marked as sparse dots.

Chapter 8

Conclusions

8.1 Summary

This thesis has focused on developing practical and cost-effective techniques based on machine vision methodologies for accurate deformation monitoring on bridge structures. Following the field practice with a commercial vision-based system, an advanced custom-developed system is proposed for bridge monitoring purposes which is effective for non-contact multi-point simultaneous sensing. A few laboratory and field monitoring tests are performed to investigate measurement accuracy and illustrate possible sensing challenges. To overcome observed limitations, a mixed system combining cameras with accelerometers is proposed to improve sensing accuracy by a sole camera.

Chapter 2 reviews approximately one hundred papers to date about vision-based displacement sensing. Video processing methods are summarised with respect to principle, relative advantages and limitations. Applications in the two most active fields: bridge deformation and cable vibration measurement are examined followed by a summary of field challenges reported in literature.

Chapter 3 reports field practice with an expensive commercial vision-based system on two long-span bridges. The purpose of these tests is to build confidence and to present existing obstacles on vision-based monitoring. Based on field experiences, a discussion about this commercial system is provided regarding installation efforts, application ranges and measurement performance.

In Chapter 4, a custom-developed vision-based system is proposed for displacement monitoring on bridge structures. Sensing accuracy in laboratory conditions and robustness to environmental effect are evaluated through three case studies. Based on analysis results, suggestions about video processing method selection are provided.

Chapter 5 proposes a practical and autonomous data fusion method to merge displacement data with the collocated acceleration for accuracy improvement. The method is based on multi-rate Kalman filter assisted with maximum likelihood

estimation. Numerical and field demonstration examples indicate the capacity for improving displacement accuracy and widening frequency bandwidth.

Chapter 6 explores vision-based system improvement through integrating collocated accelerometer measurement. This mixed system considers automatic camera shake compensation and deploys the data fusion method (in Chapter 5) for noise reduction. Through a demonstration example on a railway bridge, the mixed system is validated to offer good performance even in undesired test conditions, e.g. involving low-contrast patterns and instability of optical instrument.

Chapter 7 investigates the feasibility of vision-based systems for bridge vibration monitoring. The monitoring structure is a cable-stayed footbridge with the span length of 109 m that is easily excited by pedestrian loads. The proposed vision-based system is employed for non-contact multiple-point monitoring for bridge deck and cables. The measurement data are implemented to extract bridge modal parameters (e.g. modal frequencies and mode shapes) and to investigate modal frequency variations under varying pedestrian loads.

8.2 Contributions

The thesis is aimed at developing practical and cost-effective techniques for accurate deformation monitoring on bridge structures. Several main contributions of this thesis are presented as follows.

1. Evaluation of the state-of-art video processing methodologies for displacement measurement.

   Video processing procedures for extracting displacement information are summarised as a three-component framework, camera calibration, target tracking and structural displacement calculation. Methods for each component are presented with respect to principle, existing applications in literature, relative advantages and limitations. The working performance of these methods are investigated through three cases studies to derive measurement accuracy in laboratory conditions and sensing robustness in field testing. Some suggestions about tracking method selection are summarised for field testing.

2. A custom-developed vision-based system for bridge monitoring.

   The system supports either consumer-grade or professional cameras and incorporates four advanced video tracking methods to adapt to different test situations. It is effective for non-contact multipoint simultaneous displacement sensing for bridge components including decks and cables.

3. A practical and autonomous data fusion method to integrate displacement with acceleration for accuracy improvement.

   The main algorithm is based on multi-rate Kalman filter assisted with the maximum likelihood estimation that is implemented to determine the necessary noise parameters automatically. The method is validated on GPS data from a long-span bridge, indicating the capacity of improving displacement accuracy and widening frequency bandwidth.

4. A mixed sensing system combining a camera with accelerometers to overcome limitations of vision-based sensing.

   The system involves automatic compensation of camera shake and autonomous data fusion. Through the demonstration on a short-span railway bridge during train passages, the system is capable of achieving high accuracy even in undesired test conditions.

5. Data interpretation of vision-based measurement, i.e. ambient modal analysis and bridge influence line estimation.

   Vision-based systems are demonstrated for vibration monitoring on a footbridge. The measured data enables accurate estimation of modal frequencies and could be used to investigate variations of modal frequencies under varying pedestrian loads. Besides, displacement data collected from a railway bridge is implemented for bridge influence line estimation, indicating the potential for bridge condition evaluation.

8.3 Comments about vision-based monitoring

Vision-based monitoring follows a general procedure including three main steps, camera calibration, target tracking and structural displacement calculation. The challenging part of developing and implementing these systems lies on the selection of appropriate algorithms for each step. This study evaluates the influence of camera calibration and target tracking methods on measurement accuracy and robustness, leading to some conclusions that might be helpful to those who are interested in this topic.

8.3.1 Single-view or multiple-view

Most field applications on bridge structures rely on single-view systems for 2D deformation measurement (e.g. along bridge vertical and transverse directions) due to installation convenience. This configuration is based on the assumption that the out-of-plane motion is negligible. This can be feasible for most bridge

applications since deformation in the vertical direction is dominant compared with those in two horizontal directions.

There are also several research efforts investigating the feasibility of vision-based monitoring for 3D deformation measurement. One viable method is real-time pose estimation using single-view configuration. However, the calibration process requires a rigid body or plane near the region of interest, involving at least four non-collinear points with precisely known geometry, which is mostly provided by attaching artificial targets. The achieved accuracy might be poorer than that for 2D displacement data using offline transformation and it is impossible to recover motions along camera optical axis. Therefore, this method is recommended only when the target size is not negligible compared to the camera-to-target distance.

The second feasible way for 3D reconstruction is based on stereoscopic vision (e.g. from two or multiple views). A multi-camera arrangement provides more reliable results than a single view but the measurement quality has high dependency on the time-synchronisation of camera recordings. Currently, most validation studies are based on laboratory practices instead of field testing. One challenging aspect in the field is about configuring multiple cameras involving overlapped views about the target region on stable locations with desired angles between any two optical line of sight to the target region.

Instead of reconstructing 3D measurement, another potential for implementing the multiple-view configuration is for multi-point measurement. The distributed sensing feature for multiple point simultaneous monitoring (e.g. the application example on a footbridge in Chapter 7) is one vital advantage of vision-based monitoring while the system faces the trade-off between good sensing accuracy (or resolution) and a wide field-of-view to cover a large region of interest. A multiple-view configuration resolves this issue once the system synchronisation is achieved.

8.3.2 Tracking accuracy and robustness

Selecting proper tracking algorithms is vital to ensure a desired working performance of vision-based monitoring. In Chapter 4, three advanced tracking methods (i.e. correlation-based template matching, LK optical flow estimation and SIFT method) were evaluated in laboratory and field conditions.

Laboratory studies indicate that the tracking accuracies for the two methods (i.e. correlation-based template matching and LK optical flow estimation) are close, varying between 0.02 pixel and 0.10 pixel depending on the saliency of target patterns, while the accuracy of SIFT method is poorer in the range between 0.03 pixel and 0.20 pixel.

Although the tracking accuracy becomes slightly poorer in the field than that achieved in the laboratory, all the three tracking methods are effective for either

short-range or long-range measurement (e.g. camera-to-target distance at 710 m) with the displacement varying from several millimetres to ten centimetres. High-contrast and salient targets are always preferred for tracking and LK optical flow estimation is the most sensitive to target distinctiveness and stability.

The most challenging part is about the robustness to lighting changes in field testing. Lighting variations could occur due to cloud passing or changing weather conditions, which possibly cause some low-frequency error that is confusing without the prior knowledge of structural loading. Currently, none of the three methods could totally avoid this type of error although the SIFT method suffers less. Apart from restraining the test conditions, e.g. testing in overcast weather or selecting the sheltered target patterns, how to deal with varying lighting conditions in field testing is still an open question.

Camera and support motion induced by environment or human interventions could influence measurement accuracy as well. Currently, a common correction method is to additionally track the 'nominal' motion of a fixed target e.g. bridge towers or adjacent buildings. Recent work by Khaloo and Lattanzi (2017) proposed another promising approach for camera motion compensation through removing the averaged motion of background pixels based on dense optical flow estimation. However, all these methods depend on the condition that a static region near tracked target is available within the field of view. A proper correction method for general applications deserves further study.

8.3.3 Error in projection transformation

Observations from both laboratory and field tests indicate that although the vision-based system offers the 2D or 3D displacement data, the measurement along minor deformed direction might not be reliable. This is because the error in the camera calibration step might lead to leakage of dominant deformation to the minor deformed direction. The camera calibration step involves uncertainty about the provided geometry information, such as the given distance between two control points and the specified axis directions in a defined structural coordinate system, etc. Therefore, special attention should be given to interpret the measured displacement along the minor deformed direction (e.g. bridge longitudinal or transverse directions).

Lens distortion is commonly taken as an influential factor to sensing accuracy in literature and is properly corrected through camera calibration. Analysis results in Chapter 6.4.1 for a consumer-grade camera (i.e. GoPro 4) illustrate that considering lens distortion correction makes negligible changes on target motions in the image plane. This finding indicates that it is still feasible to perform a quick monitoring test based on a single known dimension (using the scaling factor method) near the target region without considering any lens distortion influence.

8.3.4 Practical applications

Most chapters in this thesis focus on accurate deformation measurement using vision-based systems. The systems have been validated to be viable for short-time monitoring (e.g. several hours in daytime) on road, railway bridges and footbridges. Measured data could be implemented for extracting bridge influence lines, model calibration and dynamic modal testing, etc. that could assist to solve a few practical bridge monitoring issues.

Measurement of deformation during controlled vehicle load testing helps to estimate bridge load carrying capacity. Due to increasing traffic loads and frequency as well as material deterioration, whether in-service bridges, especially aging ones, satisfy new service demand remain uncertain. Therefore, accurate assessment of bridge load carrying capacity (e.g. through field testing) becomes essential and critical for bridge owners. Non-contact vision-based systems offer a feasible way for measuring bridge displacement at a considerable cost and without influence on bridge normal operation.

When recorded at high sample rates, displacement data provide valuable information about dynamic characteristics and, hence, changes in structural condition. One potential application is to evaluate cable tensions in cable-stayed bridges. Cables are important components that transmit the major loads acting on the girder to the pylon, directly support structural weight and traffic/human loading, and control the entire internal force distribution of the deck system. Thus, the cable tension is a crucial factor in determining the overall condition assessment of a cable-stayed bridge structure. Vibration method by which the cable force is estimated from natural frequencies using a pre-determined formula (Zui et al., 1996) or finite element approach, is often used due to simplicity, rapidness and economic feature. Accelerometers are generally mounted on cable surface for vibration measurement and natural frequency identification, which require direct access to cables and also high installation efforts. Non-contact vision-based systems offer an alternative for cable vibration measurement owing the advantages e.g. quick and easy installation and features of distributed sensing (single camera for multiple point measurement simultaneously). Also, the application could be extended to monitor other structural types with slender line-shape components like transmission towers.

Although the potential of vision-based systems for ambient modal test has been demonstrated in Chapter 7, it is better not to be over-optimistic about their capacity of extracting structural dynamic properties. For road and railway bridges, traffic-induced deformations are always dominated by static and quasi-static components while the dynamic components with relatively low signal-to-noise ratios are easily contaminated by measurement noise. Thus, it is impossible to achieve comparable performance as accelerometers for capturing structural dynamics although a high

sample rate is provided for a displacement sensor (e.g. vision-based system, LVDT or GPS).

8.4 Future work

Further research topics are addressed as follows.

1. Exploring the potential of vision-based systems for long-term bridge monitoring.

   Most of bridge field tests in this research are single daytime measurement or lasting for several hours. Long-term data collection potentially offers more information about bridge conditions while the realisation still requires efforts in a few aspects.

   - Real-time system involving data acquisition and video processing. The vision-based system developed in this study records video files on site and do post-processing afterwards. Currently, the video processing package could process over 20 images per second which is sufficient for real-time monitoring on most bridge structures. The next step is to incorporate video acquisition (using cameras supporting live streaming) and video processing together for an integrated real-time system .
   - System configuration suitable for permanent or long-term running. Undesired camera motion is one important error source that should be avoided in the field. A new system configuration including stable mounts and desired shelter is necessary to replace common tripods.
   - Efficient and robust feature tracking algorithms to deal with environmental variations and tracking failure. Stable lighting condition is desired for vision-based monitoring and active illumination should be provided for night recording. Tracking failure possibly occurs during feature occlusion, appearance change (e.g. snow or rain) and lighting variations, etc. To deal with possible feature pattern changes, an automatic algorithm for feature re-localisation and update is necessary to get rid of data loss. Also, any latest feature tracking method appeared in computer vision field that could provide better robustness to environmental changes deserves validation studies for bridge monitoring.

2. Interpreting vision-based measurement to assist bridge condition evaluation.

   This work investigates the viability for ambient modal analysis and bridge influence line estimation on bridge examples. Further applications of vision-based monitoring could be explored on bridge structures such as contributing

to bridge weight-in-motion system, identifying abnormal events and damage detection/localisation, etc.

3. Drone applications.

   A proper configuration of camera set-up (including camera-to-target distance and camera orientation) is necessary to ensure a desired working performance of the system. For example, the accessible stable location for system installation in the Humber bridge application in Chapter 3 was near one bridge tower foundation, leading to a distance to the mid-span target at approximately 710 m. This configuration limits the achievable sensing resolution to millimetre level, much larger than that (sub-millimetre level) in laboratory practices. Camera systems equipped in drones have more flexibility on system configuration compared with ground-based systems, e.g. allowing closer view for better sensing resolution. However, how to overcome the minor motion of drones themselves is a challenging topic for accurate deformation monitoring.

Appendix A

Video processing methods

A detailed description about video processing methods implemented in section 4.2 is presented in this section. Four options of tracking methods are included in the package to adapt to different field testing conditions, i.e. correlation-based template matching, Lucas–Kanade optical flow estimation, scale invariant feature transform (SIFT) and edge detection for cable tracking.

A.1 Correlation-based template matching

Template matching is a tracking technique for finding an area among a source image that has the highest similarity with a template based on area intensities. To identify the matching area, the template is slid against the source image one pixel at a time (from left to right, up to down) to calculate a correlation matrix that represents the matching level against the template image. The matching location is corresponding to where the highest metric is taken within correlation matrix .

The similarity metric used to evaluate matching level is zero-mean normalised cross correlation coefficient (ZNCC). ZNCC criterion offers the most robust noise-proof performance and insensitive to the offset and linear scale in illumination lighting. The image intensity (f or g) is normalised by subtracting the mean value (f_m or g_m) and dividing by the standard deviation (Δf or Δg) as follows:

$$C_{ZNCC} = \sum_{i,j} \left[\frac{(f(i,j) - f_m)(g(i,j) - g_m)}{\Delta f \Delta g} \right] \tag{A.1}$$

The initial estimate produces a pixel-level matching location that might not be the exact position of the matching entity. In the fine search step, one can either up-sample image intensity matrix directly or up-sample correlation matrix to a higher resolution. Since the computation cost is very high for the first option, interpolation schemes are commonly preferred. The method used in this study is frequency-domain zero padding based on matrix multiplication form of discrete Fourier

transform (Guizar-Sicairos et al., 2008). The procedures to find the subpixel-level matching location are demonstrated in Fig. A.1 .

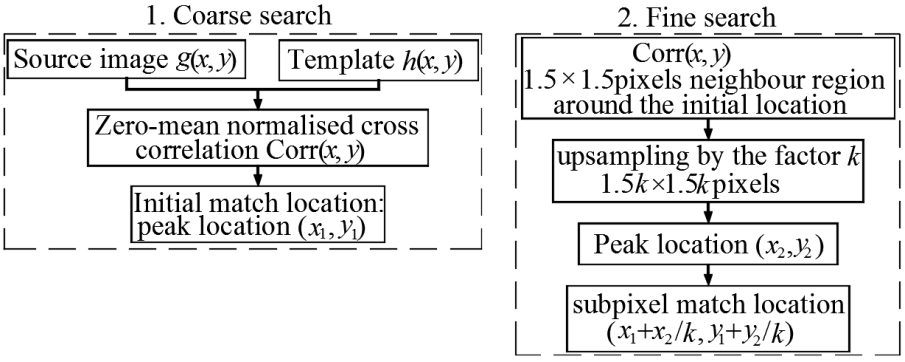


Fig. A.1 Procedures of correlation-based template matching.

Cross correlation between two signals (g and h) in time domain is equivalent to element-wise multiplication of the Fourier transform (FT) ($G(f)$) of one signal by the complex conjugate ($H^{\star}(f)$) of the FT of the other in frequency domain.

$$Corr(g,h) \longleftrightarrow G(f)H^{\star}(f) \tag{A.2}$$

Cross correlation between two up-sample signals (g_k and h_k) is equivalent to up-sample the cross correlation matrix when dealing it in frequency domain.

$$g_k \longleftrightarrow \begin{bmatrix} \mathrm{O} & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & G(f) & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} \end{bmatrix}, h_k \longleftrightarrow \begin{bmatrix} \mathrm{O} & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & H(f) & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} \end{bmatrix} \tag{A.3}$$

$$Corr(g_k,h_k) \longleftrightarrow \begin{bmatrix} \mathrm{O} & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & G(f)H^{\star}(f) & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} \end{bmatrix} \tag{A.4}$$

Thus, the refined target location could be found through implementing inverse FT to zero-padded frequency-domain correlation matrix. The computation efforts for two-dimensional (2D) inverse FT are high even the fast FT is applied. Since only the neighbouring region around the initial matching location within the correlation matrix is of interest, the interpolation process could be narrowed to the neighbouring region to enhance the efficiency. Here the matrix multiplication form of discrete inverse Fourier transform is used,

$$t_k(x,y) = W_{kN_r} T_k(x,y) W_{kN_c} \tag{A.5}$$

where $t(x,y)$ and $t_k(x,y)$ denote the 2D raw and up-sampled signals in time domain; $T(x,y)$ and $T_k(x,y)$ denote the frequency-domain signals; W_{kN_r} and W_{kN_c}

are Fourier transform matrices with the dimensions of $kN_r \times kN_r$ and $kN_c \times kN_c$, respectively.

The neighbouring region around the initial match location is interpolated by 2 first to get 0.5 pixel resolution. Then the 1.5 $\times$ 1.5 pixels neighbouring region in the correlation matrix is interpolated by a factor k (e.g. k = 100). Fig. A.2 illustrates the efficient calculation scheme for inverse Fourier transform.

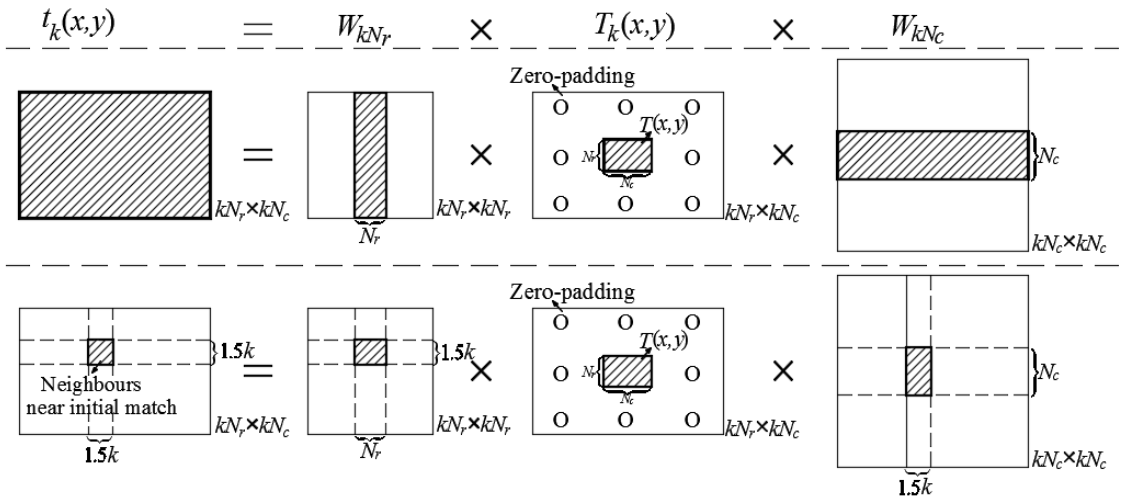


Fig. A.2 Illustration of matrix multiplication form of 2D inverse discrete Fourier transform.

A.2 Lucas–Kanade optical flow estimation

Lucas-Kanade (LK) feature tracker is a well-known method for feature tracking and sparse optical flow estimation proposed by Lucas and Kanade (1981), Tomasi and Kanade (1991), Shi et al. (1994). The procedures involves three steps, 1) detecting good feature points within the region of interest (ROI) in the first image, 2) applying LK algorithm to compute the location of each feature point in image sequences, 3) determining the ROI location after discarding outliers.

Features in computer vision field represent keypoints (or corners), edges, lines and other structures well localised in two dimensions (Szeliski, 2011). Edges are not suitable for tracking due to the 'aperture problem' except that the vibration is one dimensional translation with known direction e.g. cable tracking. Keypoints which have a local 2D structure, can be used to find a sparse set of corresponding locations in different images. Stable and salient keypoints are required from the ROI of the first frame and the popular keypoint detection method is Harris corner detector (Harris and Stephens, 1988) which is widely used in monitoring applications (Chang and Ji, 2007; Khuc and Catbas, 2017a; Yoon et al., 2016).

The LK algorithm is applied to track the motion of keypoints in image sequences based on the assumptions of brightness constancy and spatial coherence, i.e.

the projection of the same feature point keeps the same pixel intensity in image sequences and a small window around each feature point has the distortion shape following a specific warp function $W(\mathbf{x}; p)$. The goal is to minimise the dissimilarity between the two small windows including a specific keypoint in the consecutive images I and J,

$$\epsilon = \sum_{\mathbf{x}} \left[J(W(\mathbf{x}; p)) - I(\mathbf{x})\right]^2 \omega(\mathbf{x}) \tag{A.6}$$

where $\mathbf{x} = [x, y]^T$ is a column vector containing image coordinates; $W(\mathbf{x}; p)$ denotes the wrap function that maps one feature point from the initial location $\mathbf{x}$ in the first frame I to a new location $W(\mathbf{x}; p)$ in the second frame J; and $\omega(\mathbf{x})$ is weight function usually set to a constant 1. For civil monitoring applications, the warp function could be simplified to allow only planar translation as

$$W(\mathbf{x}; p) = \mathbf{x} + \mathbf{d} \tag{A.7}$$

where $\mathbf{d}$ is a column vector describing the 2D translation $[dx, dy]$. The displacement vector $\mathbf{d}$ could be obtained by minimising the expression in Eq. A.6, leading to

$$\begin{bmatrix} \sum_i I_{xi}^2 & \sum_i I_{xi} I_{yi} \\ \sum_i I_{xi} I_{yi} & \sum_i I_{yi}^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} -\sum_i I_{xi} I_{ti} \\ -\sum_i I_{yi} I_{ti} \end{bmatrix} \tag{A.8}$$

where I_x, I_y and I_t denote the spatial and temporal gradients of image intensities; i denotes the ith pixel location in a square patch (e.g. 3×3) around a feature point (x, y).

To verify the matched keypoint correspondences, RANdom SAmple Consensus (Fischler and Bolles, 1981) is implemented to discard outliers. The averaged image motion of remaining keypoints is taken as the tracking result.

A.3 Scale invariant feature transform (SIFT)

SIFT (Lowe, 2004) is a method for extracting from images distinctive features that are invariant to image scale and rotation.

The first stage is to find scale-space extrema. It is impossible to detect keypoints with different scales from the same window. Thus, the scale-space of an image is constructed that is a continuous function defined by the convolution of a variable-scale Gaussian $G(x, y, \sigma)$ with an input image $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) \circledast I(x, y) \tag{A.9}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{A.10}$$

To efficiently detect stable keypoint locations in scale space, the difference-of-Gaussian (DOG) function is computed from differences between two adjacent scales that provides a close approximation to the scale-normalised Laplacian of Gaussian:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) \circledast I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{A.11}$$

Once this DoG is generated, images are searched for local extrema over scale and space. Each sample points (one pixel) is compared with its eight neighbours in the current image as well as nine neighbours in the scale above and below (See Fig. A.3). It is a potential keypoint if it is a local extrema. It basically means that keypoint is best represented in that scale.

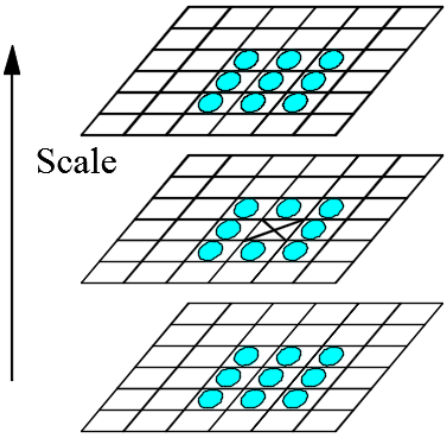


Fig. A.3 Extrema of the difference of Gaussian images detected through comparing with 26 neighbours (Lowe, 2004).

The second step is to localise qualified keypoints. The potential keypoint locations found in the first step have to be refined for more accurate estimates. Taylor series expansion (up to quadratic terms) of scale space is applied:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x} \tag{A.12}$$

where D and its derivatives are evaluated at the sample point and $\mathbf{x} = (x, y, \sigma)$ is the offset from this point. The location of the extrema $\hat{\mathbf{x}}$ is determined by

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}. \tag{A.13}$$

Poor keypoints are discarded if the contrast value is smaller than a defined threshold (e.g. 0.03 with respect to the range [0,1]) or the edge responses evaluated by the Hessian matrix are high.

In the third step, an orientation is assigned to each keypoint based on local image proprieties to achieve invariance to image rotation. A neighbourhood region

around the keypoint is taken to calculate gradient magnitudes and directions. An orientation histogram (including 36 bins covering 360 degree range) is formed for gradient orientations of sample points within the region. Peaks in the orientation histogram correspond to dominant directions of local gradients. The highest peak is detected as well as any other peak above 80% of the highest one. These peaks create keypoints with the same location and scale but different orientations and contribute significantly to the stability of matching (Lowe, 2004).

The fourth step is descriptor creation. A 16×16 neighbourhood around the keypoint is taken to compute gradient magnitude and orientation at each sample point. These samples are accumulated into eight-bin orientation histogram over 4×4 subregions, leading to a 128-element vector representation for each keypoint descriptor. In addition, the feature vector is modified (through normalisation) to reduce effects of illumination change.

Finally, keypoints between two images are matched by identifying their nearest neighbours evaluated using the Euclidean distances between keypoint descriptor vectors. The outliers in keypoint correspondences are removed based on geometric alignment using the least median of squares method (Massart et al., 1986).

A.4 Edge detection method

The cable tracking method based on edge detection is implemented in Chapter 7 for a cable-stayed footbridge monitoring under pedestrian loads. The method consists of two main steps: edge detection and motion estimation. Edge detection is aimed at determining the cable location in a small subset window while the cable motion is estimated from the distance between two extracted edges.

One of the gradient-based edge detectors, Sobel operator (Sobel, 1978) is used to detect the probable edge points (in pixel level) followed by Zernike moment operator (Ghosal and Mehrotra, 1993; Ying-Dong et al., 2005) that is applied to re-located the edge precisely. The mathematical background about these two edge detection methods are presented in this section.

A.4.1 Sobel detection

Sobel operator (Sobel, 1978) is applied first to derive pixel-level edge points. The operator uses two kernels (for horizontal and vertical directions, respectively), which are convolved with the source image to calculate derivative approximation.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \circledast Img \tag{A.14}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \circledast Img \tag{A.15}$$

$$G_{amp} = \sqrt{G_x^2 + G_y^2} \tag{A.16}$$

A threshold to the gradient magnitude (G_{amp}) is set to distinguish valid edge points. This method is fast but crude.

A.4.2 Zernike moment operator

Zernike moment operator (Ghosal and Mehrotra, 1993; Ying-Dong et al., 2005) is then applied to re-located the edge precisely from the points detected by Sobel operator. Zernike moments are constructed by mapping image data onto a set of complex polynomials (V_{nm}), which form a complete orthogonal set over the interior of a unit circle i.e. $x^2 + y^2 \leq 1$ (Ghosal and Mehrotra, 1993). Zernike moment of order n and repetition m for an image $f(x, y)$ is defined as

$$A_{nm} = \sum_x \sum_y f(x, y) V_{nm}^{\star}(\rho, \theta)(x^2 + y^2 \leq 1). \tag{A.17}$$

In (Ghosal and Mehrotra, 1993), three Zernike moments (A_{00}, A_{11} and A_{20}) are used to calculate edge parameters (Fig. A.4) of every edge point: k is the step height, L is the perpendicular distance from the centre of kernel; and the edge makes an angle of Φ with respect to x axis.

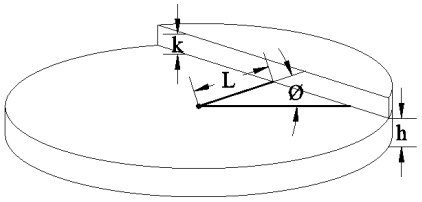


Fig. A.4 Two-dimensional step edge model.

To estimate these Zernike moments, the corresponding masks of any desired size (7 in this study) can be obtained by evaluating the associated intergral over each pixel using circular limits of integration (See Fig. A.5). The masks (Z_{00}, Z_{11}

and Z_{20}) are deduced as

$$Z_{00} = \begin{bmatrix} 0.000 & 0.029 & 0.069 & 0.081 & 0.069 & 0.029 & 0.000 \\ 0.029 & 0.082 & 0.082 & 0.082 & 0.082 & 0.082 & 0.029 \\ 0.069 & 0.082 & 0.082 & 0.082 & 0.082 & 0.082 & 0.069 \\ 0.081 & 0.082 & 0.082 & 0.082 & 0.082 & 0.082 & 0.081 \\ 0.069 & 0.082 & 0.082 & 0.082 & 0.082 & 0.082 & 0.069 \\ 0.029 & 0.082 & 0.082 & 0.082 & 0.082 & 0.082 & 0.029 \\ 0.000 & 0.029 & 0.069 & 0.081 & 0.069 & 0.029 & 0.000 \end{bmatrix} \tag{A.18}$$

$$Re(Z_{11}) = \begin{bmatrix} 0.000 & -0.015 & -0.019 & 0.000 & 0.019 & 0.015 & 0.000 \\ -0.022 & -0.047 & -0.023 & 0.000 & 0.023 & 0.047 & 0.022 \\ -0.057 & -0.047 & -0.023 & 0.000 & 0.023 & 0.047 & 0.057 \\ -0.069 & -0.047 & -0.023 & 0.000 & 0.023 & 0.047 & 0.069 \\ -0.057 & -0.047 & -0.023 & 0.000 & 0.023 & 0.047 & 0.057 \\ -0.022 & -0.047 & -0.023 & 0.000 & 0.023 & 0.047 & 0.022 \\ 0.000 & -0.015 & -0.019 & 0.000 & 0.019 & 0.015 & 0.000 \end{bmatrix} \tag{A.19}$$

$$Im(Z_{11}) = \begin{bmatrix} 0.000 & 0.022 & 0.057 & 0.069 & 0.057 & 0.022 & 0.000 \\ 0.015 & 0.047 & 0.047 & 0.047 & 0.047 & 0.047 & 0.015 \\ 0.019 & 0.023 & 0.023 & 0.023 & 0.023 & 0.023 & 0.019 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.019 & -0.023 & -0.023 & -0.023 & -0.023 & -0.023 & -0.019 \\ -0.015 & -0.047 & -0.047 & -0.047 & -0.047 & -0.047 & -0.015 \\ 0.000 & -0.022 & -0.057 & -0.069 & -0.057 & -0.022 & 0.000 \end{bmatrix} \tag{A.20}$$

$$Z_{20} = \begin{bmatrix} 0.000 & 0.022 & 0.039 & 0.040 & 0.039 & 0.022 & 0.000 \\ 0.022 & 0.027 & -0.013 & -0.026 & -0.013 & 0.027 & 0.022 \\ 0.039 & -0.013 & -0.053 & -0.066 & -0.053 & -0.013 & 0.039 \\ 0.040 & -0.026 & -0.066 & -0.079 & -0.066 & -0.026 & 0.040 \\ 0.039 & -0.013 & -0.053 & -0.066 & -0.053 & -0.013 & 0.039 \\ 0.022 & 0.027 & -0.013 & -0.026 & -0.013 & 0.027 & 0.022 \\ 0.000 & 0.022 & 0.039 & 0.040 & 0.039 & 0.022 & 0.000 \end{bmatrix} \tag{A.21}$$

The moments are estimated by convolving these masks with the image points

$$A_{nm} = Z_{nm} \circledast Img. \tag{A.22}$$

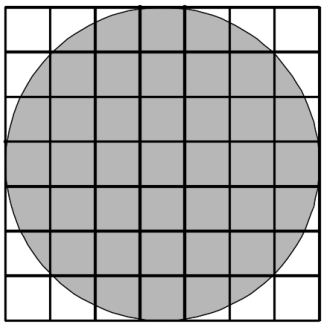


Fig. A.5 Circular kernel defined for a 7 $\times$ 7 pixel area.

The relationships between edge parameters and Zernike moments are given as below:

$$\Phi = tan^{-1}\left(\frac{Im[A_{11}]}{Re[A_{11}]}\right) \tag{A.23}$$

$$L = \frac{A_{20}}{A'_{11}} \tag{A.24}$$

$$k = \frac{3A'_{11}}{2(1-L^2)^{3/2}} \tag{A.25}$$

The calculated edge parameters are then used as criteria to remove outliers and to refine image coordinates for the remaining edge points. An edge point is identified if the detected step size k and the edge distance L follows:

$$k \leq \tau; L \geq \delta \tag{A.26}$$

where τ is a chosen threshold and δ is taken as $1/\sqrt{2}$ of the size of one pixel. Image coordinates of the qualified edge points in sub-pixel level are derived using the calculated edge parameters.

References

Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). Freak: Fast retina keypoint. In *IEEE conference on Computer vision and pattern recognition (CVPR)*, pages 510–517.

Ashkenazi, V. and Roberts, G. W. (1997). Experimental Monitoring of the Humber Bridge Using Gps. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, 120(4):177–182.

Baker, S. and Matthews, I. (2004). Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255.

Baqersad, J., Poozesh, P., Niezrecki, C., and Avitabile, P. (2017). Photogrammetry and optical methods in structural dynamics – A review. *Mechanical Systems and Signal Processing*, 86:17–34.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3):346–359.

BBC News (2015a). Forth Road Bridge to be closed until new year.

BBC News (2015b). Lorry tests as monitors installed in Forth Road Bridge first.

Beauchemin, S. S. and Barron, J. L. (1995). The computation of optical flow. *ACM Computing Surveys*, 27(3):433–466.

Bing, P., Hui-min, X., Bo-qin, X., and Fu-long, D. (2006). Performance of sub-pixel registration algorithms in digital image correlation. *Measurement Science and Technology*, 17(6):1615–1621.

Boukerroui, D., Noble, J. A., and Brady, M. (2004). On the Choice of Band-Pass Quadrature Filters. *Journal of Mathematical Imaging and Vision*, 21(1):53–80.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Sebastopol, CA.

Brownjohn, J. M., Koo, K.-Y., Scullion, A., and List, D. (2015). Operational deformations in long-span bridges. *Structure and Infrastructure Engineering*, 11(4):556–574.

Brownjohn, J. M. W., Bocian, M., Hester, D., Quattrone, A., Hudson, W., Moore, D., Goh, S., and Lim, M. S. (2016a). Footbridge system identification using wireless inertial measurement units for force and response measurements. *Journal of Sound and Vibration*, 384:339–355.

Brownjohn, J. M. W. and Botfield, T. (2009). A folded pendulum isolator for evaluating accelerometer performance. *Experimental Techniques*, 33(1):33–37.

Brownjohn, J. M. W., Hester, D., Xu, Y., Bassitt, J., and Koo, K. (2016b). Viability of optical tracking systems for monitoring deformations of a long span bridge. In *6th European Conference on Structural Control*, Sheffield, UK. European Association for the Control of Structures.

Brownjohn, J. M. W., Moyo, P., Omenzetter, P., and Lu, Y. (2003). Assessment of highway bridge upgrading by dynamic testing and finite-element model updating. *Journal of Bridge Engineering*, 8(3):162–172.

Brownjohn, J. M. W., Xu, Y., and Hester, D. (2017). Vision-Based Bridge Deformation Monitoring. *Frontiers in Built Environment*, 3:1–16.

Bülow, T. (1999). *Hypercomplex spectral signal representations for the processing and analysis of images*. PhD thesis, Christian-Albrechts-Universität zu Kiel.

Busca, G., Cigada, A., Mazzoleni, P., and Zappa, E. (2014). Vibration Monitoring of Multiple Bridge Points by Means of a Unique Vision-Based Measuring System. *Experimental Mechanics*, 54:255–271.

Caetano, E., Silva, S., and Bateira, J. (2007). Application of a vision system to the monitoring of cable structures. In *Seventh International Symposium on Cable Dynamics*, pages 225–236.

Caetano, E., Silva, S., and Bateira, J. (2011). A vision system for vibration monitoring of civil engineering structures. *Experimental Techniques*, 35(4):74–82.

Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: Binary robust independent elementary features. In *European conference on computer vision*, pages 778–792, Crete, Greece.

Carden, E. P. and Fanning, P. (2004). Vibration Based Condition Monitoring: A Review. *Structural Health Monitoring: An International Journal*, 3(4):355–377.

Carmona, R., Hwang, W.-L., and Torresani, B. (1998). *Practical Time-Frequency Analysis*, volume 9. Academic Press.

Casciati, F. and Fuggini, C. (2009). Engineering vibration monitoring by GPS: long duration records. *Earthquake Engineering and Engineering Vibration*, 8(3):459–467.

Cha, Y.-J., Chen, J., and Büyüköztürk, O. (2017). Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. *Engineering Structures*, 132:300–313.

Chan, W. S., Xu, Y. L., Ding, X. L., and Dai, W. J. (2006). An integrated GPS-accelerometer data processing technique for structural deformation monitoring. *Journal of Geodesy*, 80(12):705–719.

Chang, C. C. and Ji, Y. F. (2007). Flexible Videogrammetric Technique for Three-Dimensional Structural Vibration Measurement. *Journal of Engineering Mechanics*, 133(6):656–664.

Chang, C. C. and Xiao, X. H. (2010a). An integrated visual-inertial technique for structural displacement and velocity measurement. *Smart Structures and Systems*, 6(9):1025–1039.

Chang, C. C. and Xiao, X. H. (2010b). Three-Dimensional Structural Translation and Rotation Measurement Using Monocular Videogrammetry. *Journal of Engineering Mechanics*, 136(7):840–848.

Charalampous, E., Psimoulis, P., Guillaume, S., Spiridonakos, M., Klis, R., Bürki, B., Rothacher, M., Chatzi, E., Luchsinger, R., and Feltrin, G. (2015). Measuring sub-mm structural displacements using QDaedalus: a digital clip-on measuring system developed for total stations. *Applied Geomatics*, 7(2):91–101.

Chen, C.-C., Wu, W.-H., Tseng, H.-Z., Chen, C.-H., and Lai, G. (2015a). Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders. *Measurement*, 75:134–146.

Chen, J. G., Davis, A., Wadhwa, N., Durand, F., Freeman, W. T., and Buyukozturk, O. (2017). Video Camera-Based Vibration Measurement for Civil Infrastructure Applications. *Journal of Infrastructure Systems*, 23(3):11.

Chen, J. G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W. T., and Buyukozturk, O. (2015b). Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345:58–71.

Choi, H.-S., Cheung, J.-H., Kim, S.-H., and Ahn, J.-H. (2011). Structural dynamic displacement vision system using digital image processing. *NDT & E International*, 44(7):597–608.

Choi, I., Kim, J., and Kim, D. (2016). A Target-Less Vision-Based Displacement Sensor Based on Image Convex Hull Optimization for Measuring the Dynamic Response of Building Structures. *Sensors*, 16(12):2085.

Diamond, D., Heyns, P., and Oberholster, A. (2017). Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence. *Measurement*, 95:166–172.

Ehrhart, M. and Lienhart, W. (2015a). Development and evaluation of a long range image-based monitoring system for civil engineering structures. In Shull, P. J., editor, *Proc. SPIE Structural Health Monitoring and Inspection of Advanced Materials, Aerospace, and Civil Infrastructure*, volume 9437, page 94370K, San Diego, California, United States.

Ehrhart, M. and Lienhart, W. (2015b). Image-based dynamic deformation monitoring of civil engineering structures from long ranges. In Lam, E. Y. and Niel, K. S., editors, *Proc. SPIE*, volume 9405, page 94050J, San Francisco, California. International Society for Optics and Photonics.

Ehrhart, M. and Lienhart, W. (2015c). Monitoring of Civil Engineering Structures using a State-of-the-art Image Assisted Total Station. *Journal of Applied Geodesy*, 9(3):174–182.

Exeter Chief News (2017). Chiefs 36 Falcons 14.

Feng, D., Feng, M., Ozer, E., and Fukuda, Y. (2015a). A Vision-Based Sensor for Noncontact Structural Displacement Measurement. *Sensors*, 15(7):16557–16575.

Feng, D. and Feng, M. Q. (2015). Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads. *Journal of Bridge Engineering*, 20(12):1–12.

Feng, D. and Feng, M. Q. (2017). Experimental validation of cost-effective vision-based structural health monitoring. *Mechanical Systems and Signal Processing*, 88:199–211.

Feng, D., Scarangello, T., Feng, M. Q., and Ye, Q. (2017). Cable tension force estimate using novel noncontact vision-based sensor. *Measurement*, 99:44–52.

Feng, M. Q., Fukuda, Y., Feng, D., and Mizuta, M. (2015b). Nontarget Vision Sensor for Remote Measurement of Bridge Dynamic Response. *Journal of Bridge Engineering*, 20(12):04015023.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Fleet, D. J. and Jepson, A. D. (1990). Computation od Component Image Velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104.

Fukuda, Y., Feng, M. Q., Narita, Y., Kaneko, S., and Tanaka, T. (2013). Vision-Based Displacement Sensor for Monitoring Dynamic Response Using Robust Object Search Algorithm. *IEEE Sensors Journal*, 13(12):4725–4732.

Fukuda, Y., Feng, M. Q., and Shinozuka, M. (2010). Cost-effective vision-based system for monitoring dynamic response of civil engineering structures. *Structural Control and Health Monitoring*, 17(8):918–936.

Gautama, T. and Van Hulle, M. (2002). A phase-based approach to the estimation of the optical flow field using spatial filtering. *IEEE Transactions on Neural Networks*, 13(5):1127–1136.

Ghosal, S. and Mehrotra, R. (1993). Orthogonal moment operators for subpixel edge detection. *Pattern Recognition*, 26(2):295–306.

Greenbaum, R. J. Y., Smyth, A. W., and Chatzis, M. N. (2016). Monocular Computer Vision Method for the Experimental Study of Three-Dimensional Rocking Motion. *Journal of Engineering Mechanics*, 142(1):04015062.

Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R. (2008). Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158.

Guo, J. and Zhu, C. (2016). Dynamic displacement measurement of large-scale structures based on the Lucas–Kanade template tracking algorithm. *Mechanical Systems and Signal Processing*, 66-67:425–436.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of Fourth Alvey Vision Conference*, volume 15, Manchester, UK.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press.

Hartley, R. I. and Mundy, J. L. (1993). Relationship between photogrammmetry and computer vision. In *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision*, volume 1944, pages 92–105.

Hester, D., Brownjohn, J., Bocian, M., and Xu, Y. (2017). Low cost bridge load test: Calculating bridge displacement from acceleration for load assessment calculations. *Engineering Structures*, 143:358–374.

Ho, H.-N., Lee, J.-H., Park, Y.-S., and Lee, J.-J. (2012). A Synchronized Multipoint Vision-Based System for Displacement Measurement of Civil Infrastructures. *The Scientific World Journal*, 2012:1–9.

Jáuregui, D. V., White, K. R., Woodward, C. B., and Leitch, K. R. (2003). Non-contact Photogrammetric Measurement of Vertical Bridge Deflection. *Journal of Bridge Engineering*, 8(4):212–222.

Ji, Y. F. and Chang, C. C. (2008a). Nontarget Image-Based Technique for Small Cable Vibration Measurement. *Journal of Bridge Engineering*, 13(1):34–42.

Ji, Y. F. and Chang, C. C. (2008b). Nontarget Stereo Vision Technique for Spatiotemporal Response Measurement of Line-Like Structures. *Journal of Engineering Mechanics*, 134(6):466–474.

Jiang, R., Jáuregui, D. V., and White, K. R. (2008). Close-range photogrammetry applications in bridge measurement: Literature review. *Measurement*, 41(8):823–834.

Khaloo, A. and Lattanzi, D. (2017). Pixel-wise structural motion tracking from rectified repurposed videos. *Structural Control and Health Monitoring*, 24(11):e2009.

Khuc, T. and Catbas, F. N. (2017a). Completely contactless structural health monitoring of real-life structures using cameras and computer vision. *Structural Control and Health Monitoring*, 24(1):e1852.

Khuc, T. and Catbas, F. N. (2017b). Computer vision-based displacement and vibration monitoring without using physical target on structures. *Structure and Infrastructure Engineering*, 13(4):505–516.

Kim, J., Kim, K., and Sohn, H. (2014). Autonomous dynamic displacement estimation from data fusion of acceleration and intermittent displacement measurements. *Mechanical Systems and Signal Processing*, 42(1-2):194–205.

Kim, S. C., Kim, H. K., Lee, C. G., and Kim, S. B. (2006). A vision system for identifying structural vibration in civil engineering constructions. In *SICE-ICASE International Joint Conference*, pages 5813–5818, Bexco, Busan, Korea.

Kim, S.-W., Jeon, B.-G., Kim, N.-S., and Park, J.-C. (2013). Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge. *Structural Health Monitoring: An International Journal*, 12(5-6):440–456.

Kim, S.-W. and Kim, N.-S. (2011). Multi-point Displacement Response Measurement of Civil Infrastructures Using Digital Image Processing. *Procedia Engineering*, 14:195–203.

Kim, S.-W. and Kim, N.-S. (2013). Dynamic characteristics of suspension bridge hanger cables using digital image processing. *NDT & E International*, 59:25–33.

Koo, K. Y., Brownjohn, J. M. W., List, D. I., and Cole, R. (2013). Structural health monitoring of the Tamar suspension bridge. *Structural Control and Health Monitoring*, 20(4):609–625.

Lages Martins, L., Nunes Vicente Rebordão, J. M., and Silva Ribeiro, Á. (2014). Thermal Influence on Long-Distance Optical Measurement of Suspension Bridge Displacement. *International Journal of Thermophysics*, 35(3-4):693–711.

Lee, J.-h., Cho, S., and Sim, S.-h. (2015). Monocular Vision-based Displacement Measurement System Robust to Angle and Distance Using Homography. In *6th International Conference on Advances in Experimental Structural Engineering*, Urbana-Champaign, United States.

Lee, J. J., Cho, S., Shinozuka, M., Yun, C.-b., Lee, C.-G., and Lee, W.-T. (2006). Evaluation of Bridge Load Carrying Capacity Based on Dynamic Displacement Measurement Using Real-time Image Processing Techniques. *Steel Structures*, 6(5):377–385.

Lee, J. J. and Shinozuka, M. (2006). A vision-based system for remote sensing of bridge displacement. *NDT & E International*, 39(5):425–431.

Li, X., Ge, L., Ambikairajah, E., Rizos, C., Tamura, Y., and Yoshida, A. (2006). Full-scale structural monitoring using an integrated GPS and accelerometer system. *GPS Solutions*, 10(4):233–247.

Li, Z. and Chang, C. C. (2013). Adaptive Quanti fi cation of Noise Variance Using Subspace Technique. *Journal Of Engineering Mechanics*, 139(4):469–478.

Liao, W. Y., Chen, W. H., Ni, Y. Q., and Xia, Y. (2010). Development of a vision-based real-time displacement measurement system for Guangzhou New TV Tower. In Casciati, F. and Giordano M., editors, *Proceedings of the 5th European Workshop on Structural Health Monitoring*, pages 450–455, Sorrento, Naples, Italy.

Lin, J. and Qu, L. (2000). Feature Extraction Based on Morlet Wavelet and Its Application for Mechanical Fault Diagnosis. *Journal of Sound and Vibration*, 234(1):135–148.

Lou, P., Nassif, H., and Su, D. (2017). Impact of Heavy Freight Railcar on the Remaining Fatigue Life of Centenarian Railway Bridges. *Journal of Bridge Engineering*, 22(11):04017097.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence*, pages 674–679, Vancouver, BC, Canada. Morgan Kaufmann Publishers.

Macdonald, J. H. G., Dagless, E. L., Thomas, B. T., and Taylor, C. A. (1997). Dynamic measurements of the Second Severn Crossing. *Proceedings of the Institution of Civil Engineers - Transport*, 123(4):241–248.

Martins, L. L., Rebordão, J. M., and Ribeiro, Á. S. (2014). Optical Metrology applied to 3D displacement measurement of long-span suspension bridge dynamics. In A. Cunha, E. Caetano, P. Ribeiro, G. M., editor, *Proceedings of the 9th International Conference on Structural Dynamics*, pages 2135–2142, Porto, Portugal.

Martins, L. L., Rebordão, J. M., and Ribeiro, A. S. (2015). Structural observation of long-span suspension bridges for safety assessment: implementation of an optical displacement measurement system. *Journal of Physics: Conference Series*, 588(1):012004.

Massart, D. L., Kaufman, L., Rousseeuw, P. J., and Leroy, A. (1986). Least median of squares: a robust method for outlier and model error detection in regression and calibration. *Analytica Chimica Acta*, 187(C):171–179.

McCormick, N., Waterfall, P., and Owens, A. (2014). Optical imaging for low-cost structural measurements. *Proceedings of the Institution of Civil Engineers - Bridge Engineering*, 167(1):33–42.

Min, J.-H., Gelo, N. J., and Jo, H. (2016). Real-time image processing for non-contact monitoring of dynamic displacements using smartphone technologies. In Lynch, J. P., editor, *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, page 98031B, Las Vegas, United States. SPIE.

Moeslund, G. and Granum, E. (2001). A survey of computer vision-based human motion capture.

Moreu, F., Jo, H., Li, J., Kim, R. E., Cho, S., Kimmle, A., Scola, S., Le, H., Spencer, B. F., and LaFave, J. M. (2015). Dynamic Assessment of Timber Railroad Bridges Using Displacements. *Journal of Bridge Engineering*, 20(10):04014114.

Moschas, F., Psimoulis, P. A., and Stiros, S. C. (2013). GPS / RTS data fusion to overcome signal deficiencies in certain bridge dynamic monitoring projects. *Smart Structures and Systems*, 12(3):251–269.

Moschas, F. and Stiros, S. (2011). Measurement of the dynamic displacements and of the modal frequencies of a short-span pedestrian bridge using GPS and an accelerometer. *Engineering Structures*, 33(1):10–17.

Moschas, F. and Stiros, S. (2013). Noise characteristics of high-frequency, short-duration GPS records from analysis of identical, collocated instruments. *Measurement: Journal of the International Measurement Confederation*, 46(4):1488–1506.

Moschas, F. and Stiros, S. (2014). PLL bandwidth and noise in 100 Hz GPS measurements. *GPS Solutions*, 19(2):173–185.

Nickitopoulou, A., Protopsalti, K., and Stiros, S. (2006). Monitoring dynamic and quasi-static deformations of large flexible engineering structures with GPS: Accuracy, limitations and promises. *Engineering Structures*, 28(10):1471–1482.

Oh, B. K., Hwang, J. W., Kim, Y., Cho, T., and Park, H. S. (2015). Vision-based system identification technique for building structures using a motion capture system. *Journal of Sound and Vibration*, 356:72–85.

Ojio, T., Carey, C. H., OBrien, E. J., Doherty, C., and Taylor, S. E. (2016). Contactless Bridge Weigh-in-Motion. *Journal of Bridge Engineering*, 21(7):04016032.

Olaszek, P. (1999). Investigation of the dynamic characteristic of bridge structures using a computer vision method. *Measurement*, 25(3):227–236.

Pan, B., Qian, K., Xie, H., and Asundi, A. (2009). Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. *Measurement Science and Technology*, 20(6):062001.

Park, C., Ho, H.-N., Jo, B.-W., and Lee, J.-J. (2014). An Efficient Vision-Based Three-Dimensional Motion Measurement System for Civil Infra-Structures. *Experimental Techniques*, 40(2):1–9.

Park, J.-W., Moon, D.-S., Yoon, H., Gomez, F., Spencer Jr., B. F., and Kim, J. R. (2018). Visual-inertial displacement sensing using data fusion of vision-based displacement with acceleration. *Structural Control and Health Monitoring*, 25(3):e2122.

Park, S., Park, H., Kim, J., and Adeli, H. (2015). 3D displacement measurement model for health monitoring of structures using a motion capture system. *Measurement*, 59:352–362.

Peeters, B. and De Roeck, G. (1999). Reference-Based Stochastic Subspace Identification for Output-Only Modal Analysis. *Mechanical Systems and Signal Processing*, 13(6):855–878.

Potter, K. D. and Setchell, C. (2014). Positional measurement of a feature within an image.

Prendergast, L. and Gavin, K. (2014). A review of bridge scour monitoring techniques. *Journal of Rock Mechanics and Geotechnical Engineering*, 6(2):138–149.

Prendergast, L., Hester, D., Gavin, K., and O'Sullivan, J. (2013). An investigation of the changes in the natural frequency of a pile affected by scour. *Journal of Sound and Vibration*, 332(25):6685–6702.

Psimoulis, P. a. and Stiros, S. C. (2012). A supervised learning computer-based algorithm to derive the amplitude of oscillations of structures using noisy GPS and Robotic Theodolites (RTS) records. *Computers and Structures*, 92-93:337–348.

Rahbari, R., Niu, J., Brownjohn, J. M. W., and Koo, K. Y. (2015). Structural identification of Humber Bridge for performance prognosis. *Smart Structures and Systems*, 15(3):665–682.

Ribeiro, D., Calçada, R., Ferreira, J., and Martins, T. (2014). Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system. *Engineering Structures*, 75:164–180.

Roberts, G. W., Meng, X., and Dodson, A. H. (2004). Integrating a Global Positioning System and Accelerometers to Monitor the Deflection of Bridges. *Journal of Surveying Engineering*, 130(2):65–72.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2564–2571. IEEE.

Santos, C. A., Batista, J. P., and Costa, C. O. (2009). A Non-Contact Measurement System for Monitoring the Displacements of Long Deck Suspension Bridges. In *The 16th International Road Federation World Road Meeting*, pages 1–7, Lisbon, Portugal.

Santos, C. A., Costa, C. O., and Batista, J. (2016). A vision-based system for measuring the displacements of large structures: Simultaneous adaptive calibration and full motion estimation. *Mechanical Systems and Signal Processing*, 72-73:678–694.

Schreier, H., Orteu, J.-J., and Sutton, M. A. (2009). *Image Correlation for Shape, Motion and Deformation Measurements*. Springer US, Boston, MA.

Schreier, H. W. (2004). Advances in Light Microscope Stereo Vision. *Experimental Mechanics*, 44(3):278–288.

Shan, B., Zheng, S., and Ou, J. (2015). Free vibration monitoring experiment of a stayed-cable model based on stereovision. *Measurement*, 76:228–239.

Shi, J., Tomasi, C., Jianbo Shi, and Tomasi (1994). Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, pages 593–600. IEEE Comput. Soc. Press.

Sims, C. (1999). Matlab optimization software.

Sładek, J., Ostrowska, K., Kohut, P., Holak, K., Gąska, A., and Uhl, T. (2013). Development of a vision based deflection measurement system and its accuracy assessment. *Measurement*, 46(3):1237–1249.

Smyth, A. and Wu, M. (2007). Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. *Mechanical Systems and Signal Processing*, 21(2):706–723.

Sobel, I. (1978). Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer graphics and image processing*, 8(1):127–135.

Sohn, H., Farrar, C. R., Hemez, F., and Czarnecki, J. (2002). A Review of Structural Health Monitoring Literature 1996 – 2001. In *Third World Conference on Structural Control*, Como, Italy. Los Alamos National Laboratory.

Song, Y.-Z., Bowen, C. R., Kim, A. H., Nassehi, A., Padget, J., and Gathercole, N. (2014). Virtual visual sensors and their application in structural health monitoring. *Structural Health Monitoring: An International Journal*, 13(3):251–264.

Stephen, G. A., Brownjohn, J. M. W., and Taylor, C. A. (1993). Measurements of static and dynamic displacement from visual monitoring of the Humber Bridge. *Engineering Structures*, 15(3):197–208.

Sun, D., Roth, S., and Black, M. J. (2010). Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2439.

Sutton, M., Yan, J., Tiwari, V., Schreier, H., and Orteu, J. (2008). The effect of out-of-plane motion on 2D and 3D digital image correlation measurements. *Optics and Lasers in Engineering*, 46(10):746–757.

Szeliski, R. (2011). *Computer Vision: Algorithms and Applications*, volume 5 of *Texts in Computer Science*. Springer Science & Business Media, London.

Tomasi, C. and Kanade, T. (1991). Detection and Tracking of Point Features. *International Journal of Computer Vision*.

Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. (2013). Phase-based video motion processing. *ACM Transactions on Graphics*, 32(4):1.

Wahbeh, A. M., Caffrey, J. P., and Masri, S. F. (2003). A vision-based approach for the direct measurement of displacements in vibrating systems. *Smart Materials and Structures*, 12(5):785–794.

Wang, H., Mao, J. X., Huang, J. H., and Li, A. Q. (2016). Modal Identification of Sutong Cable-Stayed Bridge during Typhoon Haikui Using Wavelet Transform Method. *Journal of Performance of Constructed Facilities*, 30(5):1–11.

Wang, N., O'Malley, C., Ellingwood, B. R., and Zureick, A.-H. (2011). Bridge Rating Using System Reliability Assessment. I: Assessment and Verification by Load Testing. *Journal of Bridge Engineering*, 16(6):854–862.

Welch, P. D. (1967). The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73.

Westgate, R., Koo, K. Y., Brownjohn, J., and List, D. (2014). Suspension bridge response due to extreme vehicle loads. *Structure and Infrastructure Engineering*, 10(6):821–833.

Wu, L. and Casciati, F. (2014). Local positioning systems versus structural monitoring: a review. *Structural Control and Health Monitoring*, 21(9):1209–1221.

Wu, L.-J., Casciati, F., and Casciati, S. (2014). Dynamic testing of a laboratory model via vision-based sensing. *Engineering Structures*, 60:113–125.

Xu, Y., Brownjohn, J., Hester, D., and Koo, K. Y. (2016). Dynamic displacement measurement of a long span bridge using vision-based system. In *8th European Workshop On Structural Health Monitoring (EWSHM 2016)*, Bilbao, Spain.

Xu, Y., Brownjohn, J., and Kong, D. (2018). A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Structural Control and Health Monitoring*, page e2155.

Xu, Y. and Brownjohn, J. M. W. (2018a). Review of machine-vision based methodologies for displacement measurement in civil structures. *Journal of Civil Structural Health Monitoring*, 8(1):91–110.

Xu, Y. and Brownjohn, J. M. W. (2018b). Vision-based systems for structural deformation measurement: case studies. *Proceedings of the Institution of Civil Engineers - Structures and Buildings*, pages 1–45.

Xu, Y., Brownjohn, J. M. W., Hester, D., and Koo, K. Y. (2017). Long-span bridges: Enhanced data fusion of GPS displacement and deck accelerations. *Engineering Structures*, 147:639–651.

Yan, B. and Miyamoto, A. (2006). A comparative study of modal parameter identification based on wavelet and Hilbert-Huang transforms. *Computer-Aided Civil and Infrastructure Engineering*, 21(1):9–23.

Yang, Y., Dorn, C., Mancini, T., Talken, Z., Kenyon, G., Farrar, C., and Mascareñas, D. (2017). Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mechanical Systems and Signal Processing*, 85:567–590.

Ye, X., Ni, Y., Wai, T., Wong, K., Zhang, X., and Xu, F. (2013). A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification. *Smart Structures and Systems*, 12(3_4):363–379.

Ye, X. W., Dong, C. Z., and Liu, T. (2016). A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications. *Journal of Sensors*, 2016:1–10.

Yi, J.-H., Kim, J.-H., Jeong, W. M., and Chae, J.-W. (2013a). Field evaluation of optical-based three-dimensional dynamic motion measurement system with multiple targets for a floating structure. *Ocean Engineering*, 62:140–151.

Yi, T. H., Li, H. N., and Gu, M. (2013b). Experimental assessment of high-rate GPS receivers for deformation monitoring of bridge. *Measurement*, 46(1):420–432.

Yi, T. H., Li, H. N., and Gu, M. (2013c). Wavelet based multi-step filtering method for bridge health monitoring using GPS and accelerometer. *Smart Structures and Systems*, 11(4):331–348.

Ying-Dong, Q., Cheng-Song, C., San-Ben, C., and Jin-Quan, L. (2005). A fast subpixel edge detection method using Sobel-Zernike moments operator. *Image and Vision Computing*, 23(1):11–17.

Yoneyama, S., Kitagawa, A., Iwata, S., Tani, K., and Kikuta, H. (2007). Bridge deflection measurement using digital image correlation. *Experimental Techniques*, 31(1):34–40.

Yoneyama, S. and Ueda, H. (2012). Bridge Deflection Measurement Using Digital Image Correlation with Camera Movement Correction. *Materials Transactions*, 53(2):285–290.

Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M., and Spencer, B. F. (2016). Target-free approach for vision-based structural system identification using consumer-grade cameras. *Structural Control and Health Monitoring*, 23(12):1405–1416.

Yuen, K.-V., Hoi, K.-I., and Mok, K.-M. (2007). Selection of noise parameters for Kalman filter. *Earthquake Engineering and Engineering Vibration*, 6(1):49–56.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

Zhao, X., Ri, K., and Wang, N. (2017). Experimental Verification for Cable Force Estimation Using Handheld Shooting of Smartphones. *Journal of Sensors*, 2017:1–13.

Zhou, H., Zheng, J., Xie, Z., Lu, L., Ni, Y., and Ko, J. (2017). Temperature effects on vision measurement system in long-term continuous monitoring of displacement. *Renewable Energy*.

Zui, H., Shinke, T., and Namita, Y. (1996). Practical Formulas for Estimation of Cable Tension By Vibration Method. *Journal of Structural Engineering*, 122(6):651–656.